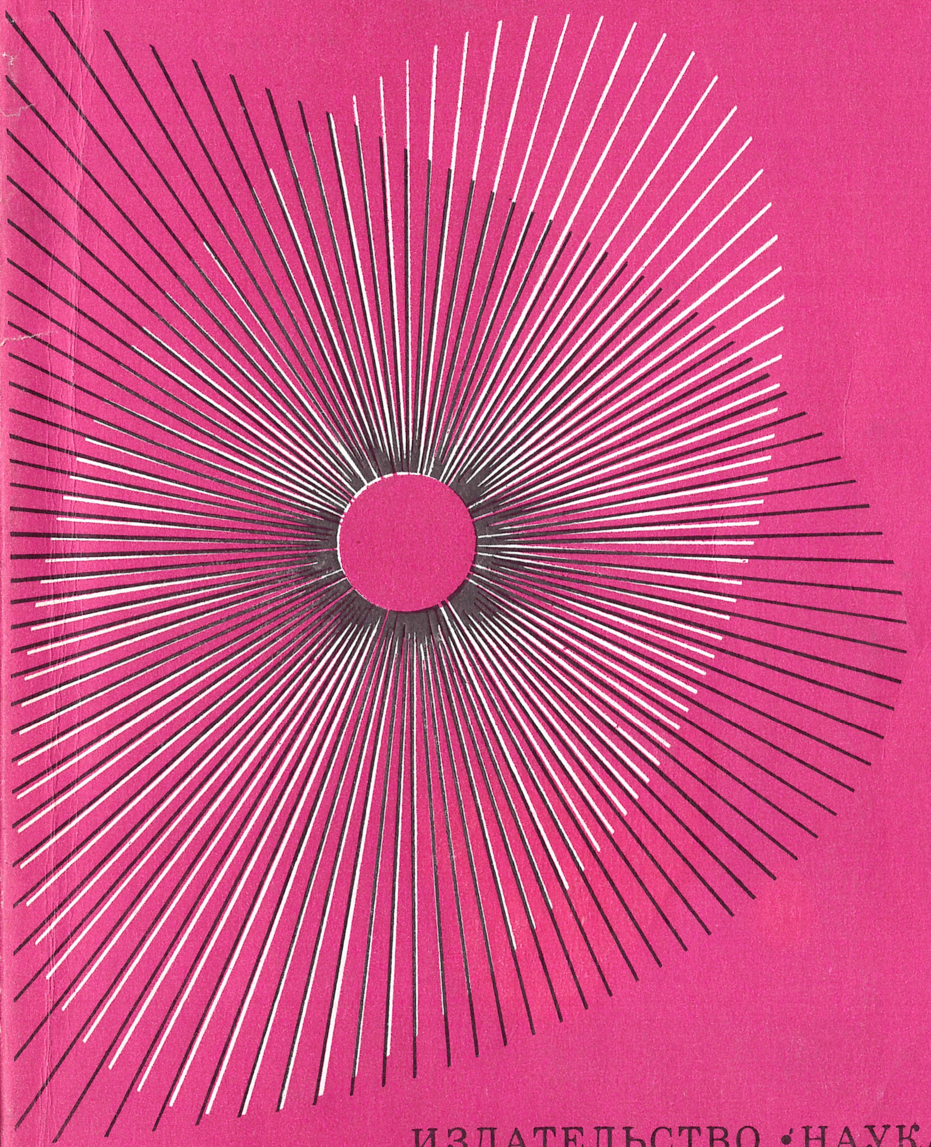
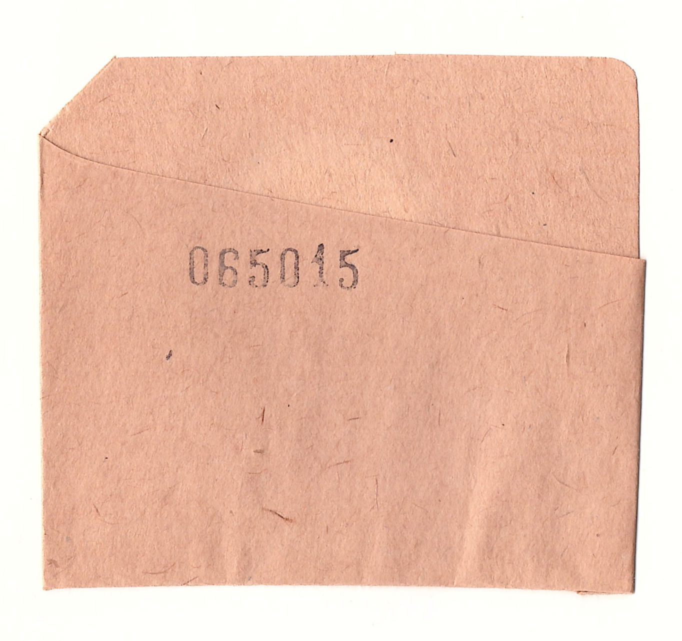
Распознавание образов



ИЗДАТЕЛЬСТВО «НАУКА»


065015

АКАДЕМИЯ НАУК СССР
ИНСТИТУТ ПРОБЛЕМ ПЕРЕДАЧИ ИНФОРМАЦИИ

Распознавание образов

ТЕОРИЯ И ПРИЛОЖЕНИЯ



ИЗДАТЕЛЬСТВО «НАУКА»
МОСКВА 1977

УДК 621.391

Сборник посвящен методам и алгоритмам решения задач прогнозирования и их применению в медицине и геологии, а также автоматическому распознаванию речи.

Обсуждаются проблемы общения человека с машиной: анализ слитной речи, иерархические структуры интонационного, слогового, лексического и синтаксического представлений, синтез речи. Приводятся алгоритмы оценивания параметров речи и идентификации сигналов, искаженных нестационарными помехами.

Рассчитан на специалистов в области распознавания образов и смежных областях.

Ответственный редактор
доктор технических наук
И. Т. ТУРБОВИЧ

Зам. ответственного редактора
кандидат технических наук
В. С. ФАЙН

Р $\frac{30501-495}{055(02)-1977}$ 828—77

Предисловие

В предлагаемый сборник включены главным образом работы, относящиеся к двум интенсивно развивающимся направлениям в распознавании образов: многомерному прогнозированию и автоматическому пониманию слитной речи.

Задача прогнозирования понимается здесь, как задача оценки состояния или положения некоторого физического объекта на основе данных, косвенно характеризующих это состояние. Особенность принятого в публикуемых работах подхода к проблеме состоит в том, что доступные сведения о влиянии указанных переменных на оцениваемый показатель, во-первых, существенно одномерны и, во-вторых, как правило, носят качественный характер.

В сборнике помещены как теоретические, так и прикладные работы этого направления. В них метод прогноза исследуется для различных способов задания прогнозируемой величины, в частности путем разбиения экспериментальной выборки на однородные и упорядоченные по этой величине классы и путем представления прогнозируемой величины дискретным рядом числовых значений. Исследуется и возможность применения аппарата проверки статистических гипотез для создания метода выявления редуцируемости используемых моделей прогнозирования.

Практическое использование этих исследований иллюстрируется работами по выявлению связи между тяжестью состояния больного и его физиологическими показателями и по прогнозированию естественного режима нефтяного пласта по совокупности геолого-геофизических характеристик.

Перерастание традиционного распознавания речевых образов в новое направление — автоматическое понимание слитной речи — привело к существенному пересмотру методов представления речи.

Слитная речь представляется последовательностью слогов и отдельных звуков, идентификация которых требует усовершен-

ствования техники анализа речи, разработанной для опознания изолированно произносимых команд. Результаты идентификации объединяются в иерархически организованной системе интонационного, слогового, лексического и синтаксического представления речи. В работах по теории и практике анализа и синтеза речи в сборнике освещается состояние проблемы и приводятся результаты ряда оригинальных исследований в этой области. Большое значение имеет и проблема представления искаженных сигналов. Одно из перспективных направлений в этой области, затрагиваемое в сборнике, состоит в идентификации параметров экспоненциальных сигналов, искаженных нестационарными помехами.

Сборник завершается работой общего характера, в которой предлагается и обосновывается новая («вариативная») модель в распознавании образов, позволяющая, в частности, использовать в этой области некоторые результаты теоретико-информационных исследований.

В. Г. Гитис

Об одном классе задач прогнозирования

В задачах прогнозирования требуется найти связь между прогнозируемым показателем r, который обычно отождествляется с состоянием некоторого физического объекта, и переменными $\{x_i\}$, $i = 1, \ldots, I$, определяющими это состояние. В более формальной постановке требуется в рамках заданной математической модели $f(\mathbf{x}, \boldsymbol{\alpha})$ связи между показателем r и вектором переменных $\mathbf{x}$ найти такую оценку параметров $\boldsymbol{\alpha}$, при которой $r \approx f(\mathbf{x}, \boldsymbol{\alpha})$.

Существует ряд задач прогнозирования, в которых показатель r не поддается объективному измерению, а задается с помощью экспертных оценок. Такие задачи встречаются в области медицинского и технического прогнозирования [1—3], в экономике, социологии и психологии [4, 5]. Часто в этих задачах экспертная оценка показателя r может быть введена лишь с помощью разбиения всех реализаций экспериментальной выборки на однородные и упорядоченные по этому показателю классы $\omega_1, \omega_2, \ldots, \omega_Q$. В этом случае меру соответствия прогнозируемого показателя r и прогноза $f(x, \alpha)$ можно ввести, потребовав, чтобы решающее правило θ, найденное по учебной выборке с использованием модели $f(x, \alpha)$, позволяло в некотором смысле наилучшим образом классифицировать и упорядочить элементы контрольной выборки в соответствии с экспертными оценками показателя r. Если для вектора параметров $\boldsymbol{\alpha}^*$ упорядоченность элементов контрольной выборки по решающему правилу достаточно близка к упорядоченности этих же элементов по экспертным оценкам, то будем считать, что прогноз $f(x, \alpha)$ с точностью до монотонной функции соответствует прогнозируемому показателю r.

Определим критерий качества решающего правила. Введем функцию потерь $S(n_q \mid k)$, $n_q = 1, \ldots, N_q$, $q = 1, \ldots, Q$, $k = 1, \ldots, Q$, которая характеризует потери, возникающие при отнесении реализации $x_{n_q} \in \boldsymbol{\omega}_q$ к классу $\boldsymbol{\omega}_k$ по решающему правилу θ. Качество решающего правила можно оценить эмпирическим средним риском

$$R = \frac{1}{N} \sum_{q=1}^{Q} \sum_{n=1}^{N_q} S(n_q \mid k), \tag{1}$$

где $N = \sum_{q=1}^{Q} N_q$ — общее число элементов выборки.

Заметим, что отнесение реализации x_{n_q} к классу ω_k вызывает нарушение отношения порядка реализации x_{n_q} для $|q-k|$ классов. Например, если реализация x_{n_q} отнесена решающим правилом θ к классу ω_{q+p} или ω_{q-p}, то в том и другом случае имеет место нарушение отношения порядка реализации x_{n_q} для p классов: соответственно $\omega_{q+1}, \omega_{q+2}, \ldots, \omega_{q+p}$ или $\omega_{q-1}, \omega_{q-2}, \ldots, \omega_{q-p}$. Поэтому особенностью функции потерь $S(n_q|k)$ по сравнению с задачами классификации, в которых нет необходимости в сохранении отношения порядка, является требование быть невозрастающей функцией от k при $k < q$ и неубывающей функцией при $k > q$, т. е.

$$S(n_q|k) = \begin{cases} S_{1n_q}(k) & \text{при } k < q, \\ 0 & \text{при } k = q, \\ S_{2n_q}(k) & \text{при } k > q. \end{cases} \tag{2}$$

Так как функции $S_{1n_q}(k)$ и $S_{2n_q}(k)$ являются монотонными, то для любого k их можно представить в виде сумм, состоящих из неотрицательных элементов:

$$S_{1n_q}(k) = \sum_{p=1}^{q-1} s_{n_q p},$$

$$S_{2n_q}(k) = \sum_{p=q+1}^{k} s_{n_q p}, \qquad s_{n_q p} \geqslant 0, \tag{3}$$

где элементы $s_{n_q p}$ характеризуют потери, возникающие при нарушении отношения порядка реализации x_{n_q} для класса ω_p.

При выборе функции потерь необходимо учесть то обстоятельство, что ошибки классификации по найденному решающему правилу могут быть вызваны не только «ошибками алгоритма», связанными с неточностью математической модели, с неточностью измерений переменных $\{x_i\}$, с представительностью учебной выборки и т. д., но ошибочной классификацией самого эксперта. Ошибки эксперта могут быть обусловлены как недостатком информации при разбиении выборки на классы, так и тем, что в подобных задачах само понятие «класс» часто является весьма условным, что в общем случае делает безошибочную классификацию выборки экспертом принципиально невозможной.

Рассмотрим два примера выбора функции потерь. Пусть для каждой реализации $x_{n_q} \in \omega_q$ функция потерь линейно зависит от числа классов, для которых произошло нарушение порядка, т. е. $S(n_q|k) = s_{n_q}|q-k|$. При этом элементы $s_{n_q p}$ зависят только от номера реализации, вызвавшей нарушение порядка, т. е. $s_{n_q p} \equiv s_{n_q}$. Подобная функция потерь обычно вводится в тех

случаях, когда эксперт классифицирует и упорядочивает реализации выборки с различной степенью уверенности и предполагает, что потери $s_{n_q p}$ при нарушении отношения порядка данной реализации одинаковы для любого класса $p \neq q$. Во втором примере для всех реализаций $\mathbf{x}_{n_q} \in \boldsymbol{\omega}_q$ положим $S(n_q \mid k) = 0$ при $k \in \in \{q-1, q, q+1\}$ и $S(n_q \mid k) = s_{n_q} \mid q - k - 1 \mid$ при $k \notin \{q-1, q, q+1\}$. Этот способ задания функции потерь соответствует довольно типичной ситуации, когда эксперт не в состоянии дать четкую классификацию выборки и вынужден вводить так называемую слабую упорядоченность классов. При этом эксперт исходит из предположения, что при разбиении им выборки на классы в любой класс ω_q могли попасть неизвестные ему реализации соседних классов ω_{q-1} и $\boldsymbol{\omega}_{q+1}$.

Рассмотрим алгоритм нахождения решающего правила, минимизирующего эмпирический средний риск (1).

Построение решающего правила θ можно представить в виде двухэтапной процедуры: нахождение функции $y = f(\mathbf{x}, \hat{\boldsymbol{\alpha}})$ из некоторого класса функций, задаваемого моделью связи прогнозируемого показателя r с вектором переменных $\mathbf{x}$, и выбор на числовой оси y порогов D_q, задающих границы классов. Для любого заданного вектора параметров $\hat{\boldsymbol{\alpha}}$ пороги D_q могут находиться путем минимизации (1). Найденное значение условного минимума эмпирического риска (1) примем в качестве меры соответствия прогноза $y = f(\mathbf{x}, \hat{\boldsymbol{\alpha}})$ и прогнозируемого показателя r. Безусловный минимум эмпирического среднего риска (1) может быть найден варьированием $f(\mathbf{x}, \boldsymbol{\alpha})$ по параметрам $\boldsymbol{\alpha}$ с определением оптимальных порогов D_q для каждого $\boldsymbol{\alpha}$.

При заданном векторе параметров $\boldsymbol{\alpha}$ решающее правило можно определить с помощью $Q - 1$ порогов следующим образом: $y_{n_q} \in \in \boldsymbol{\omega}_1$ при $y_{n_q} < D_1$, $y_{n_q} \in \boldsymbol{\omega}_2$ при $D_1 \leqslant y_{n_q} < D_2, \ldots, y_{n_q} \in \in \boldsymbol{\omega}_Q$ при $D_{Q-1} \leqslant y_{n_q}$. Такой способ введения решающего правила требует выполнения условия упорядоченного расположения порогов на числовой оси y, т. е. требуется

$$D_k \leqslant D_l \quad \text{при} \quad k < l. \tag{4}$$

С учетом (2) — (4) функцию потерь $S(n_q \mid k)$ можно представить в следующем виде:

$$S(n_q \mid k) = \sum_{p=1}^{Q-1} [s_{n_q p} L(y_{n_q} < D_p) + s_{n_q p+1} L(y_{n_q} \geqslant D_p)], \tag{5}$$

где

$$L(y_{n_q} < D_p) = \begin{cases} 0 \text{ при } y_{n_q} \geqslant D_p \text{ для всех } & p = 1, \ldots, Q, \\ 0 \text{ при } y_{n_q} < D_p \text{ для } & p \geqslant q, \\ 1 \text{ при } y_{n_q} < D_p \text{ для } & p < q; \end{cases} \tag{6}$$

$$L(y_{n_q} \geqslant D_p) = \begin{cases} 0 \text{ при } y_{n_q} < D_p \text{ для всех } & p = 1, \ldots, Q, \\ 0 \text{ при } y_{n_q} \geqslant D_p \text{ для } & p < q, \\ 1 \text{ при } y_{n_q} \geqslant D_p \text{ для } & p \geqslant q. \end{cases} \quad (7)$$

С учетом (5) — (7) функционал (1) может быть переписан в следующем виде:

$$R = \frac{1}{N} \sum_{q=1}^{Q} \sum_{n=1}^{N_q} \sum_{p=1}^{Q-1} [s_{n_q p} L(y_{n_q} < D_p) + s_{n_q p+1} L(y_{n_q} \geqslant D_p)] =$$

$$= \frac{1}{N} \sum_{p=1}^{Q-1} \Big[\sum_{q=p+1}^{Q} \sum_{n=1}^{N_q} s_{n_q p} L(y_{n_q} < D_p) + \sum_{q=1}^{p} \sum_{n=1}^{N_q} s_{n_q p+1} L(y_{n_q} \geqslant D_p) \Big]. \quad (8)$$

При нахождении условного минимума среднего риска (8) по порогам $\{D_p\}$ для любого заданного вектора параметров α требуется, чтобы найденные пороги удовлетворяли условию (4).

Покажем, что при $s_{n_q p} \equiv s_{n_q}$ выражение (8) допускает независимую минимизацию по каждому из порогов. Обозначим через d_p^* такое значение порога d, что

$$R_p(d_p^*) = \min_d R_p(d), \quad (9)$$

где

$$R_p(d) = \sum_{q=p+1}^{Q} \sum_{n=1}^{N_q} s_{n_q} L(y_{n_q} < d) + \sum_{q=1}^{Q} \sum_{n=1}^{N_q} s_{n_q} L(y_{n_q} \geqslant d). \quad (10)$$

Рассмотрим взаимное расположение порога d_k^* и порогов d_{k+m}^*, $m = 1, \ldots, Q-k-1$. При $d_k^* \leqslant d_{k+m}^*$ для всех m условие (4) для порога d_k^* выполняется и принимается $D_k = d_k^*$.

Покажем, что если хотя бы для одного m имеет место $d_k^* > d_{k+m}^*$, то коррекция $D_k = \min\limits_m \{d_{k+m}^*\}$ сохраняет минимальное значение функционала (8).

Подсчитаем значение

$$R_k(d_{k+m}^*) = \sum_{q=k+1}^{Q} \sum_{n=1}^{N_q} s_{n_q} L(y_{n_q} < d_{k+m}^*) +$$

$$+ \sum_{q=1}^{k} \sum_{n=1}^{N_q} s_{n_q} L(y_{n_q} \geqslant d_{k+m}^*) =$$

$$= \sum_{q=k+m+1}^{Q} \sum_{n=1}^{N_q} s_{n_q} L(y_{n_q} < d_{k+m}^*) +$$

$$+\sum_{q=1}^{k+m}\sum_{n=1}^{N_q} s_{n_q} L(y_{n_q} \geqslant d^*_{k+m}) +$$

$$+\sum_{q=k+1}^{k+m}\sum_{n=1}^{N_q} s_{n_q} L(y_{n_q} < d^*_{k+m}) - \tag{11}$$

$$-\sum_{q=k+1}^{k+m}\sum_{n=1}^{N_q} s_{n_q} L(y_{n_q} \geqslant d^*_{k+m}) = R_{k+m}(d^*_{k+m}) +$$

$$+\sum_{q=k+1}^{k+m}\sum_{n=1}^{N_q} s_{n_q} [L(y_{n_q} < d^*_{k+m}) - L(y_{n_q} \geqslant d^*_{k+m})].$$

Из условия

$$R_{k+m}(d^*_{k+m}+\varepsilon) \geqslant R_{k+m}(d^*_{k+m}) \quad \text{при} \quad \varepsilon > 0, \tag{12}$$

которое имеет место в силу того, что порог d^*_{k+m} определяется согласно (9) для $p = k + m$, и условия

$$\sum_{q=k+1}^{k+m}\sum_{n=1}^{N_q} s_{n_q} [L(y_{n_q} < d^*_{k+m}+\varepsilon) - L(y_{n_q} \geqslant d^*_{k+m}+\varepsilon)] \geqslant$$

$$\geqslant \sum_{q=k+1}^{k+m}\sum_{n=1}^{N_q} s_{n_q} [L(y_{n_q} < d^*_{k+m}) - L(y_{n_q} \geqslant d^*_{k+m})] \quad \text{при} \quad \varepsilon > 0 \tag{13}$$

следует

$$R_k(d^*_{k+m}+\varepsilon) \geqslant R_k(d^*_{k+m}), \quad \varepsilon > 0. \tag{14}$$

Пусть $d^*_k = d^*_{k+m} + \varepsilon$. При этом $R_k(d^*_k) = R_k(d^*_{k+m}+\varepsilon) \leqslant \leqslant R_k(d^*_{k+m})$, так как порог d^*_k является точкой минимума функционала (9) для $p = k$. Следовательно, при $d^*_k > d^*_{k+m}$ согласно (14) возможно лишь $R_k(d^*_k) = R_k(d^*_{k+m}+\varepsilon) = R_k(d^*_{k+m})$.

Доказанное верно для любого $m = 1, \ldots, Q - k - 1$. Следовательно, если порог d^*_k находится правее нескольких порогов d^*_{k+m}, то очевидно, что коррекция $D_k = \min_m \{d^*_{k+m}\}$ не изменяет значения функционала (9) для $p = k$.

Поскольку с учетом (8) и (10)

$$R = \frac{1}{N}\sum_{p=1}^{Q-1} R_p(D_p), \tag{15}$$

то нахождение порогов согласно (9) с последующей их коррекцией соответствует нахождению условного минимума среднего риска при выполнении ограничения (4).

Таким образом, нахождение порогов $\{D_p\}$ при заданном векторе параметров α сводится к последовательному решению задач

одномерного поиска минимума (9) для $p = 1, \ldots, Q - 1$. Это значительно упрощает вычисление меры соответствия прогноза $f(\mathbf{x}, \alpha)$ прогнозируемому показателю r.

Алгоритм минимизации эмпирического среднего риска состоит из двух основных блоков: поисковой процедуры, определяющей выбор вектора параметров α, и достаточно простой процедуры вычисления условного минимума функционала (8) для заданного вектора α с использованием соотношения (9). Коррекция порогов требуется только на этапе выдачи окончательных результатов. Функционал (8) является разрывным и в общем случае многоэкстремальным по α. Для поиска вектора α, соответствующего точке глобального минимума функционала, может быть использована одна из процедур глобальной оптимизации многопараметрических функций методом случайного поиска.

Рассмотрим способ задания функции потерь, соответствующий введению слабой упорядоченности классов. Пусть для всех реализаций $\mathbf{x}_{n_q} \in \omega_q$, $q = 1, \ldots, Q$, функция потерь $S(n_q \mid k) = 0$ при $k \in \{q-1;\ q,\ q+1\}$ и $S(n_q \mid k) = s_{n_q} \mid q - k - 1 \mid$ при $k \notin \{q-1, q, q+1\}$. Так как в этом случае эмпирический средний риск (1) не изменяется при нарушении отношения порядка для соседних классов, то при заданном векторе параметров α целесообразно отказаться от поиска границ для разделения пар классов ω_1 и ω_2, ω_2 и $\omega_3, \ldots, \omega_{Q-1}$ и ω_Q. Это приводит к уменьшению числа порогов $\{D_q\}$ на единицу. Решающее правило при заданном векторе α имеет вид $y_{n_q} \in \omega_1 \cup \omega_2$ при $y_{n_q} < D_1$, $y_{n_q} \in \omega_2 \cup \omega_3$ при $D_1 \leqslant y_{n_q} < D_2, \ldots, y_{n_q} \in \omega_{Q-1} \cup \omega_Q$ при $D_{Q-2} \leqslant y_{n_q}$.

Вычисление меры соответствия прогноза $f(\mathbf{x}, \alpha)$ и прогнозируемого показателя r, как и в предыдущем случае, сводится к процедурам одномерного поиска. Но при этом выражение (10) заменяется на

$$R'_p(d) = \sum_{q=p+2}^{Q} \sum_{n=1}^{N_q} s_{n_q} L(y_{n_q} < d) + \sum_{q=1}^{p} \sum_{n=1}^{N_q} s_{n_q} L(y_{n_q} \geqslant d), \qquad (16)$$

в котором при поиске порога d_p^* исключается класс ω_{p+1}. В остальном алгоритм минимизации среднего риска не изменяется.

Значительное упрощение алгоритма отыскания минимума среднего риска (1) может иметь место при выполнении следующих предположений: 1) прогноз $f(\mathbf{x}, \alpha)$ является линейным по параметрам α, т. е. $f(\mathbf{x}, \alpha) = \alpha^T \varphi(\mathbf{x})$; 2) существует такой вектор α^*, для которого $R'_p(d) = 0$, $p = 1, \ldots, Q-2$. Это означает, что существует вектор α^*, для которого все проекции векторов, соединяющих точки различных классов ω_{q+k} и ω_q, больше нуля.

При выполнении сделанных допущений задача поиска вектора α^*, минимизирующего (1), сводится к нахождению одного из ре-

шений совместной системы линейных неравенств

$$\alpha^T [\varphi(x_{m_{q+k}}) - \varphi(x_{n_q})] > 0,$$
$$q = 1, \ldots, Q-2, \quad k = 2, \ldots, Q-q, \ m = 1, \ldots, N_{q+k},$$
$$n = 1, \ldots, N_q, \qquad (17)$$

которое может быть получено с помощью итерационных алгоритмов, детально разработанных в распознавании образов [6].

ЛИТЕРАТУРА

1. *В. И. Бунэ, Б. А. Борисов, В. Г. Гитис, Г. И. Рейснер, И. Т. Турбович, Е. Ф. Юрков.* О методе выявления связи максимально возможной магнитуды землетрясений с тектоническими параметрами района.— ДАН СССР, 1974, т. 214, № 3.
2. *Р. С. Виницкая, И. Т. Турбович, Е. Ф. Юрков, А. А. Адамян, М. Я. Авруцкий, Ю. Д. Кашин.* Прогноз дыхательной недостаточности после операции на легких по данным дооперационных исследований.— Экспериментальная хирургия и анестезиология, 1973, № 5.
3. *Р. С. Виницкая, В. С. Нагорнов, В. Я. Темников.* Применение одномерных нелинейных преобразований для прогноза тяжести послеоперационного состояния больных туберкулезом.— Сб. «Нелинейные и линейные методы в распознавании образов». М., «Наука», 1975.
4. *С. А. Айвазян.* Об опыте применения экспертно-статистического метода построения неизвестной целевой функции.— Сб. «Многомерный статистический анализ в социально-экономических исследованиях». М., «Наука», 1974.
5. Сб. «Статистическое измерение качественных характеристик». М., «Статистика», 1972.
6. *В. Н. Вапник, А. Я. Червоненкис.* Теория распознавания образов. М., «Наука», 1974.

Е. Ф. Юрков, В. С. Нагорнов

Модифицированный метод наименьших квадратов при решении задач прогноза с дискретной прогнозируемой величиной

Известен целый ряд задач прогнозирования, в которых значения прогнозируемой величины ограничены конечным (дискретным) числом значений. Возможный путь решения таких задач состоит в нахождении функционального соответствия (прогнозирующей функции), устанавливающего приближенную зависимость между дискретной прогнозируемой величиной и значениями прогнозирующих величин.

В качестве прогнозирующих функций будем рассматривать функции вида

$$f(\theta, Y) = \sum_{i=1}^{P} \theta_i \varphi_i(Y), \qquad (1)$$

где $\theta = (\theta_1, \ldots, \theta_p)^T$ — вектор параметров, подлежащих определению; Y — вектор прогнозирующих переменных; $\varphi_i(Y)$ — заданные функции векторного аргумента, причем $\varphi_1(Y) = 1$ (т. е. θ_1 является свободным членом).

Задачу нахождения прогнозирующей функции будем понимать как задачу выбора вектора, обеспечивающего наилучшее в некотором смысле совпадение между дискретной прогнозируемой величиной r и ее прогнозом $f(\theta, Y)$.

Известным методом нахождения прогнозирующей функции является метод наименьших квадратов [1], который может быть применен также и для случая, когда прогнозируемая величина является дискретной. Другой подход к решению рассматриваемых задач изложен в [2]. Согласно методу наименьших квадратов, искомый вектор $\hat{\theta}$ в (1) находится по экспериментальной выборке из условия минимума суммы квадратов S^2:

$$S^2 = \sum_{n=1}^{N} (r_n - f(\theta, Y_n))^2, \qquad (2)$$

где r_n и Y_n — значения прогнозируемой величины и вектора прогнозирующих переменных в n-м эксперименте; N — число элементов в выборке.

Специфика рассматриваемых задач состоит в том, что дискретная прогнозируемая величина r разбивает всю выборку на классы, число которых равно числу различных дискретных значений прогнозируемой величины. Поэтому, наряду с желанием минимизировать сумму квадратов (2), естественное желание состоит в том, чтобы прогноз $f(\theta, Y_n)$ для класса с прогнозируемым значением r концентрировался вокруг именно этого значения, а не около какого-нибудь другого. Рассматривая $f = f(\theta, Y)$ и r как случайные величины, последнее требование можно записать в виде условия

$$M(f/r) = r, \qquad (3)$$

где $M(f/r)$ — математическое ожидание величины f при заданном r. В других терминах $M(f/r)$ есть регрессия величины f на r. Условие (3) означает, что регрессия $M(f/r)$ должна быть линейной.

В реальных ситуациях, однако, мы всегда имеем дело с конечной выборкой. Поэтому теоретическую линейную регрессию $M(f/r)$ следует заменить выборочной линейной регрессией $l(r)$, а условие (3) заменить условием

$$l(r) = r. \qquad (4)$$

Учитывая сказанное, сформулируем критерий оптимальности для рассматриваемой здесь задачи. Будем считать оптимальным такой вектор θ в (1), который минимизирует сумму квадратов (2) при условии (4). Уточним условие (4). Так как уравнение линейной регрессии f на r имеет вид $l(r) = a + br$, то условие (4) распадается на два условия:

$$a = 0, \quad b = 1, \tag{5}$$

где, как известно,

$$a = \bar{f} - b\bar{r}, \qquad b = \sum_n (r_n - \bar{r})(f_n - \bar{f}) \Big/ \sum_n (r_n - \bar{r})^2 \tag{6}$$

и

$$\bar{r} = (1/N)\sum_n r_n, \qquad f_n = f(\theta, Y_n), \qquad \bar{f} = (1/N)\sum_n f_n.$$

В свою очередь условие (5) с учетом (6) можно переписать в виде

$$\sum_n (f_n - r_n) = 0, \qquad \sum_n r_n (f_n - r_n) = 0. \tag{7}$$

Как известно, минимизацию функционала (2) при условиях (7) можно заменить минимизацией безусловного функционала F:

$$F = \sum_n (f_n - r_n)^2 - 2\lambda \sum_n r_n (f_n - r_n) - 2\mu \sum_n (f_n - r_n), \tag{8}$$

где λ и μ — неопределенные множители Лагранжа.

Полученный при этом вектор θ и будет искомым вектором, минимизирующим выражение (2) при ограничении (4).

Обозначив $x_{ni} = \varphi_i(Y_n)$, $\bar{x}_i = (1/N)\sum_n \varphi_i(Y_n)$, выпишем необходимые условия минимума F:

$$\partial F/\partial\theta_j = 2\sum_n (f_n - r_n) x_{nj} - 2\lambda \sum_n r_n x_{nj} - 2\mu N \bar{x}_j = 0, \; j = 1, \ldots, p,$$

$$\partial F/\partial\lambda = 2\sum_n r_n (f_n - r_n) = 0, \tag{9}$$

$$\partial F/\partial\mu = 2\sum_n (f_n - r_n) = 0.$$

Перепишем (9) в матричной форме

$$(X^T X)\,\theta = \varkappa X^T R + \mu N \bar{X}, \qquad R^T X\theta = R^T R, \qquad \bar{r} = \bar{X}^T\theta, \tag{10}$$

где $X = (x_{ni})$ — матрица размера $N \times P$, $\bar{X} = (\bar{x}_1, \ldots, \bar{x}_P)^T$ и $R = (r_1, \ldots, r_N)^T$ — векторы размерности P и N, $\varkappa = \lambda + 1$.

Предполагая, что матрица $(X^T X)^{-1}$ существует, и умножая ее слева на первую формулу (10), получим

$$\theta = \varkappa (X^T X)^{-1} X^T R + \mu N (X^T X)^{-1} \bar{X}. \tag{11}$$

С учетом обозначений

$$\hat{\theta} = (X^T X)^{-1} X^T R, \quad e = N (X^T X)^{-1} \bar{X} \qquad (12), \quad (13)$$

перепишем (11):

$$\theta = \varkappa\hat{\theta} + \mu e. \tag{14}$$

Отметим, что вектор $\hat{\theta}$ минимизирует сумму квадратов (2), если снять ограничение (4) (это следует из (11) при $\lambda = \mu = 0$). Другими словами, $\hat{\theta}$ есть вектор, к которому приводит метод наименьших квадратов.

Вектор e имеет вид:

$$e = (e_1, \ldots, e_p)^T = (1, 0, 0, \ldots, 0)^T. \tag{15}$$

Чтобы убедиться в этом, обозначим $X^TX = (g_{ij})$, $(X^TX)^{-1} = (g^{ij})$. Заметим, что $g_{1j} = N\bar{x}_j$, так как $x_{n1} = \varphi_1(Y_n) = 1$. Элемент g^{ij} обратной матрицы $(X^TX)^{-1}$ можно записать как

$$g^{ij} = G_{ij}/|X^TX|,$$

где G_{ij} — алгебраическое дополнение элемента g_{ij} прямой матрицы X^TX. Компонента e_i вектора e равна

$$e_i = N\sum_j g^{ij}\bar{x}_j = \sum_j g^{ij}g_{1j} = \frac{1}{|X^TX|}\sum_j G_{ij}g_{1j} = \begin{cases} 1, & i = 1; \\ 0, & i \neq 1. \end{cases}$$

Отсюда следует (15).

В скалярной записи выражение (14) с учетом (15) принимает следующий вид:

$$\theta_1 = \varkappa\hat{\theta}_1 + \mu, \qquad \theta_i = \varkappa\hat{\theta}_i, \qquad i = 2, \ldots, P. \tag{16}$$

Формулы (16) дают простую связь компонент искомого вектора θ с компонентами вектора $\hat{\theta}$, полученного по методу наименьших квадратов.

Аналогично получим связь прогнозирующей функции $[f_n = \sum_i \theta_i x_{ni}]$ с функцией $[\hat{f}_n = \sum_i \hat{\theta}_i x_{ni}]$, найденной по методу наименьших квадратов:

$$f_n = \theta_1 + \sum_{i=2}^{P} \theta_i x_{ni} = \varkappa \sum_{i=1}^{P} \hat{\theta}_i x_{ni} + \mu = \varkappa\hat{f}_n + \mu. \tag{17}$$

Неопределенные множители $\varkappa$ и μ можно найти, подставив (14) во вторую и третью формулы (10) $\varkappa = (R^TR - N\bar{r}^2)/(R^TX\hat{\theta} - N\bar{r}^2)$, $\mu = \bar{r} - \varkappa\bar{X}^T\hat{\theta}$. Обозначив

$$S_r^2 = \sum_n (r_n - \bar{r})^2 = R^TR - N\bar{r}^2, \tag{18}$$

$$\hat{S}^2 = \sum_n \left(r_n - \sum_i \hat{\theta}_i x_{ni}\right)^2 = R^TR - (X\hat{\theta})^T X\hat{\theta} \tag{19}$$

и учитывая известные в методе наименьших квадратов соотношения $R^T X\hat{\theta} = (X\hat{\theta})^T X\hat{\theta}$, $\bar{r} = \bar{X}^T\hat{\theta}$, получим окончательно

$$\varkappa = S_r^2/(S_r^2 - \hat{S}^2), \qquad \mu = -\bar{r}\hat{S}^2/(S_r^2 - \hat{S}^2). \tag{20}$$

Из (20) непосредственно следует, что $1/\varkappa$ является квадратом (выборочного) множественного коэффициента корреляции между величинами r и x_i $(i = 1, \ldots, P)$.

Отметим, что так как величины f и $\hat{f}$ связаны линейным соотношением (17), а коэффициент корреляции инвариантен относительно линейных преобразований коррелирующих величин, то коэффициенты корреляции r с f и r с $\hat{f}$ равны друг другу. Из метода наименьших квадратов известно [1], что коэффициент корреляции между r и $\hat{f}$ достигает максимума по всем прогнозирующим функциям вида (1) и равен множественному коэффициенту корреляции. Отсюда следует, что в классе функций (1) рассматриваемый здесь метод прогнозирования также максимизирует коэффициент корреляции между r и f, и величина этого коэффициента корреляции равна $1/\sqrt{\varkappa}$.

Другую интерпретацию для $\varkappa$ можно получить, рассмотрев линейную регрессию $l(r)$ величины $\hat{f}$ на r:

$$\hat{l}(r) = \hat{a} + \hat{b}r, \tag{21}$$

где коэффициенты линейной регрессии $\hat{a}$ и $\hat{b}$ в соответствии с (6) равны:

$$\hat{a} = \bar{\hat{f}} - \hat{b}\bar{r}, \quad \hat{b} = \sum_n (r_n - \bar{r})(\hat{f}_n - \bar{\hat{f}})/\sum_n (r_n - \bar{r})^2, \tag{22}$$

$$\bar{\hat{f}} = \frac{1}{N}\sum_n \hat{f}_n.$$

Подставляя в (6) величины $f_n - \bar{f} = \varkappa(\hat{f}_n - \bar{\hat{f}})$ и $\bar{f} = \varkappa\bar{\hat{f}} + \mu$ и учитывая условие (5), получим соотношение для $\varkappa$ и μ:

$$-\frac{\mu}{\varkappa} = \bar{\hat{f}} - \frac{1}{\varkappa}\bar{r}, \qquad \frac{1}{\varkappa} = \sum_n (r_n - \bar{r})(\hat{f}_n - \bar{\hat{f}})/\sum_n (r_n - \bar{r})^2. \tag{23}$$

Сравнение (22) и (23) показывает, что $\hat{a} = -\mu/\varkappa$ и $\hat{b} = 1/\varkappa$. Таким образом, величины $-\mu/\varkappa$ и $1/\varkappa$ являются коэффициентами линейной регрессии $\hat{l}(r)$ (21).

Получим выражение для суммы квадратов (2), соответствующее найденному вектору θ. С учетом второй формулы (7) имеем

$$S^2 = \sum_n (f_n - r_n)^2 = \sum_n f_n^2 - \sum_n r_n^2.$$

С учетом (17) и первой формулы (23)

$$\sum_n f_n^2 = \varkappa \sum_n \left(\hat{f}_n - \bar{\hat{f}} + \frac{1}{\varkappa}\bar{r}\right)^2 = \varkappa^2 \left(\sum_n \left(\hat{f}_n - \bar{\hat{f}}\right)^2 + \frac{N}{\varkappa^2}\bar{r}^2\right).$$

Используя обозначения (18), (19) и соотношение метода наименьших квадратов $\sum_n (\hat{f} - \bar{\hat{f}})^2 = S_r^2 - \hat{S}^2$, получаем $\sum_n f_n^2 = \varkappa^2 (S_r^2 - \hat{S}^2) + N\bar{r}^2$.

Окончательно, с учетом первой формулы (20) и (18), имеем

$$S^2 = \frac{(S_r^2)^2}{(S_r^2 - \hat{S}^2)^2}(S_r^2 - \hat{S}^2) - S_r^2 = \frac{S_r^2}{S_r^2 - \hat{S}^2}\hat{S}^2 = \varkappa \hat{S}^2. \qquad (24)$$

Полученные результаты показывают, что предложенный метод решения задачи прогнозирования с дискретной прогнозируемой величиной сводится к небольшой модификации метода наименьших квадратов. Это позволяет при нахождении параметров прогнозирующей функции использовать стандартные программы метода наименьших квадратов с последующим пересчетом параметров по простым формулам (16), (20) и (24).

Представляет интерес сравнение предложенного метода прогнозирования с методом наименьших квадратов. Ясно, однако, что для сравнения нельзя использовать критерий суммы квадратов (2), так как без ограничения (4) этот критерий не учитывает дискретной специфики рассматриваемой задачи. Коэффициент корреляции между прогнозируемой величиной и ее прогнозом также нельзя использовать, так как было показано, что он одинаков для двух сравниваемых методов. Очевидно, для сравнения следует взять критерий, который бы не был связан с «внутренними» критериями сравниваемых методов.

Такой критерий легко получить, воспользовавшись следующим обстоятельством. Прогноз по формуле (1) дает значение непрерывной величины, которую следует превратить в дискретную. Естественный способ дискретизации состоит в том, что значение непрерывной величины f_n округляется до ближайшего допустимого значения $[f_n]$ дискретной прогнозируемой величины. Критерием качества прогнозирования может служить число ошибок прогнозирования, равное числу случаев, для которых $[f_n] \neq r_n$.

Экспериментальное сравнение изложенного здесь метода с методом наименьших квадратов производилось на материале задачи с дискретной прогнозируемой величиной, описанной в [3].

Число ошибок в первом случае оказалось на 18% меньше, чем во втором случае. Заметим, что уменьшение числа ошибок связано не с увеличением числа свободных параметров алгоритма прогнозирования, а с более правильным выбором критерия качества в случае дискретной прогнозируемой величины.

ЛИТЕРАТУРА

1. *С. Р. Рао.* Линейные статистические методы и их применения. М., «Наука», 1968.
2. *В. Г. Гитис.* Об одном классе задач прогнозирования.— Наст. сб.
3. *И. Т. Турбович, Р. С. Виницкая, В. Г. Гитис, С. Г. Ерамян, В. С. Нагорнов, Н. Н. Сунгурян, Е. Ф. Юрков.* Анализ связи между тяжестью состояния больного и его физиологическими показателями.— Наст. сб.

Е. Ф. Юрков

Статистический анализ модели прогнозирования, использующей одномерные нелинейные преобразования

Прогноз скалярной величины r, согласно методу прогнозирования, основанному на использовании одномерных нелинейных преобразований, ищут по критерию наименьших квадратов в виде [1]

$$f(X) = z\left(\sum_{i=1}^{I} \varphi_i(x_i)\right), \quad (1)$$

где $X = (x_1, \ldots, x_I)$ — вектор прогнозирующих переменных; z и φ_i — одномерные монотонные функции, подлежащие определению по выборке экспериментальных данных.

При решении конкретных задач прогнозирования может, однако, оказаться, что связь между прогнозируемой величиной и переменными x_i описывается более простой зависимостью, чем (1). Возможна, например, аддитивная или даже просто линейная зависимость. Поэтому при решении практических задач может возникнуть вопрос, обладает ли модель (1) реальным преимуществом перед более простыми моделями прогнозирования.

В настоящей статье на материале задач медицинского и сейсмического прогнозирования [1—3] производится сравнение модели (1) с ее частными случаями: H_1, H_2, H_3:

$$H_1 : f_1(X) = a_0 + \sum_{i=1}^{I} a_i x_i, \quad (2)$$

$$H_2 : f_2(X) = \sum_{i=1}^{I} \varphi_i(x_i), \quad (3)$$

$$H_3 : f_3(X) = z\left(a_0 + \sum_{i=1}^{I} a_i x_i\right), \quad (4)$$

где z и φ_i — функции того же класса, что и в (1); a_i — параметры линейной формы.

В предположении о том, что вид функций z и φ_i в (1) определяется некоторым числом свободных параметров θ_k, выражения (2) — (4) можно получить, налагая определенное число связей на эти параметры. Действительно, пусть, например, функция $z = z(y)$ задается значениями $z(y_k) = \theta_k$ в нескольких точках y_k, а в остальных точках получается линейной интерполяцией соседних значений $z(y_k)$. Тем самым вид функции z полностью определяется набором параметров θ_k. Таким образом, условие $z(y) = y$, приводящее (1) к (3), сводится к заданию связей вида $\theta_k = y_k$. Аналогично получим, что выражения (2) и (4) также являются результатом выполнения некоторого числа соотношений, связывающих между собой свободные параметры в (1).

Указанное обстоятельство позволяет задачу о выборе того или иного прогнозирующего выражения свести к задаче проверки заданных соотношений, связывающих параметры θ_k. Как известно, последняя задача может быть решена методами проверки статистических гипотез [4—7].

В соответствии с логикой проверки статистических гипотез выражения (2) — (4) рассматриваются как гипотезы H_1, H_2, H_3, которые следует либо принять, либо отклонить. Может оказаться, что все гипотезы (2) — (4) отклоняются, тогда прогнозирование на основе выражения (1) будем считать оправданным. В другом, крайнем случае, когда ни одна из гипотез (2) — (4) не отклоняется, целесообразно принять самую простую из них, т. е. гипотезу H_1 о линейном прогнозе.

Отметим, что задача о сокращении числа переменных x_i в выражениях (1) — (4) также может быть решена в рамках проверки статистических гипотез. В этом случае соотношения, связывающие свободные параметры, должны привести к гипотезе $\varphi_i(x_i) \equiv 0$ для некоторой группы переменных x_i. Если гипотеза принимается, то указанная группа переменных не является существенной для прогноза. Если гипотеза отклоняется, то это значит, что в указанной группе имеются существенные переменные, которые нельзя отбросить [7].

В настоящей работе в качестве статистического критерия, на основании которого принимается либо отклоняется гипотеза, используется критерий дисперсионного отношения (F-критерий).

Дисперсинное отношение, как известно, имеет вид

$$F = \frac{(S_H^2 - S^2)(N - p)}{S^2 m_H}, \tag{5}$$

где $S^2 = \min \sum_{n=1}^{N} (r_n - f(X_n))^2$ — безусловная сумма квадратов (в отсутствии связей на параметры); p — суммарное число независимых

параметров в модели прогнозирования; $S_H^2 = \min \sum_{n=1}^{N} (r_n - f_H(X_n))^2$ — условная сумма квадратов (при наличии связей, налагаемых гипотезой H); m_H — число связей, налагаемых гипотезой H; N — число выборочных реализаций.

При этом гипотеза H отклоняется, если F превосходит значение критической точки $F_{кр}$, взятой из статистических таблиц, в противном случае H принимается. Применительно к рассматриваемому случаю гипотеза H — одно из выражений (2) — (4), S^2 и p относятся к выражению (1), S_H^2 — к выражениям (2) — (4), $m_H = p - p_H$, p_H — число независимых параметров в выражениях (2) — (4).

Как известно, использование F-критерия требует выполнения следующих допущений:

а) предполагается, что конкретный набор значений величин r и X является результатом одного из N независимых испытаний, причем для n-го испытания прогнозируемая величина r_n рассматривается как случайная величина, имеющая нормальное распределение с математическим ожиданием $Mr_n = f(X_n)$ и неизвестной дисперсией, одинаковой для всех n;

б) предполагается, что как зависимость функции $f(X)$ от параметров, так и связи, налагаемые гипотезой являются линейными по параметрам.

В практических задачах вряд ли следует ожидать, что эти допущения выполняются строго. Так, например, распределение величины r_n (допущение а)) может отклоняться от нормального. Однако, как отмечено в [5], это обстоятельство не является достаточной причиной для отказа от применения статистического критерия, поскольку вследствие действия центральной предельной теоремы даже значительное отклонение от нормальности приводит лишь к небольшому смещению F-отношения.

Первая часть допущения б) о линейности $f(X)$ по параметрам также не выполняется. Тем не менее, если в определенной области пространства параметров степень нелинейности по параметрам незначительна, то и погрешность статистического вывода будет незначительной [6]. Вторая часть допущения б) о линейности связей, налагаемых гипотезой, выполняется при условии, что одномерные функции z и φ_i линейны по параметрам.

Контроль нелинейности. Рассмотрим способ контроля нелинейности функции по параметрам при проверке гипотезы вида (2) — (4). Для удобства дальнейшего изложения функцию $f(X)$ будем обозначать как $f(\theta, X)$, где $\theta = (\theta_1, \ldots, \theta_p)$ — p-мерный вектор параметров, подлежащий определению. Степень отклонения функции $f(\theta, X)$ от функции, линейной по параметрам θ, можно оценить с помощью разложения $f(\theta, X)$ в ряд Тейлора по θ. Для фиксированного вектора X_n разложение $f(\theta, X_n)$

в точке $\hat{\theta}$, учитывающее только линейные члены, имеет вид

$$f_{л}(\theta, X_n) = f(\hat{\theta}, X_n) + df(\theta, X_n), \quad (6)$$

где

$$df(\theta, X_n) = \sum_{k=1}^{p} \frac{\partial f(\theta, X_n)}{\partial \theta_k}\Bigg|_{\hat{\theta}} (\theta_k - \hat{\theta}_k) \quad (7)$$

является дифференциалом по θ для $f(\theta, X_n)$.

Рассмотрим две суммы квадратов:

$$S^2(\theta) = \sum_{n=1}^{N} (r_n - f(\theta, X_n))^2, \quad (8)$$

$$S^2_{л}(\theta) = \sum_{n=1}^{N} (r_n - f_{л}(\theta, X_n))^2. \quad (9)$$

Если бы функция $f(\theta, X)$ была линейна по θ, то $S^2(\theta) = S^2_{л}(\theta)$. В случае нелинейной зависимости величина

$$\mu(\theta) = |S^2(\theta) - S^2_{л}(\theta)| \quad (10)$$

может служить мерой линейности функции $f(\theta, X)$ по θ в области пространства параметров, включающей две точки $\hat{\theta}$ и θ [7].

При проверке гипотезы в качестве $\hat{\theta}$ выберем точку, минимизирующую безусловную сумму квадратов, а в качестве θ — точку, минимизирующую условную сумму квадратов при гипотезе H. В этом случае $\mu(\theta)$ будет мерой нелинейности функции $f(\theta, X)$ по θ при проверке гипотезы H. Если $\mu(\theta)$ мало, то $S^2(\theta) \approx \approx S^2_{л}(\theta)$, и любую из этих величин допустимо использовать в качестве S^2_H в F-отношении; в противном случае использование $S^2(\theta)$ или $S^2_H(\theta)$ в F-отношении недопустимо.

Уточним выражение (6) для случая, когда $f(X)$ имеет вид (1). Записав (1) в виде сложной функции $f(X) = z(y(X))$, $y(X) = \sum_{i=1}^{I} \varphi_i(x_i)$ и используя свойство инвариантности дифференциала сложной функции, представим сложный дифференциал (7) в виде суммы двух частных дифференциалов

$$df(\theta, X_n) = d_z f + d_y f, \quad (11)$$

где

$$d_z f = dz = z(\hat{y}(X_n)) - \hat{z}(\hat{y}(X_n)), \quad (12)$$

$$d_y f = \frac{\partial z}{\partial y}\Bigg|_{\hat{\theta}} dy = \frac{\partial \hat{z}}{\partial y}\Bigg|_{\hat{y}(X_n)} (y(X_n) - \hat{y}(X_n)). \quad (13)$$

В (12) и (13) функции $\hat{z}$ и $\hat{y}$ соответствуют вектору $\hat{\theta}$, минимизирующему безусловную сумму квадратов, а z и y — вектору θ, минимизирующему условную сумму квадратов для гипотезы H. Подставив (11) — (13) в (6), получим окончательно

$$f_{л}(\theta, X_n) = z(\hat{y}(X_n)) + \left.\frac{\partial \hat{z}}{\partial y}\right|_{\hat{y}(X_n)} (y(X_n) - \hat{y}(X_n)). \quad (14)$$

Алгоритм нахождения функций, входящих в выражение (14), приведен в [1].

Экспериментальное сравнение моделей при решении задач медицинского и сейсмического прогнозирования. Модель (1) использовалась для решения задачи прогнозирования тяжести состояния больного после операции на легких по дооперационным показателям и задачи прогнозирования максимальных энергий землетрясений на территории Кавказа [1—3].

Результаты проверки гипотез H_1, H_2, H_3, отвечающих выражениям (2) — (4), даются в таблице. Критические значения $F_{кр}$ приведены для однопроцентного уровня значимости (т. е. допускается, что в одном проценте случаев вывод об отклонении модели вида (2) — (4) является ошибочным).

Во всех случаях значение меры нелинейности (10) составляло не более 5% от величины $S^2(\theta)$, что свидетельствует о допустимости применения F-критерия в рассматриваемых задачах.

Гипотеза	Медицинская задача. $S^2 = 830$; $N = 128$; $p = 25$				Сейсмическая задача. $S^2 = 89$; $N = 91$; $p = 20$			
	S^2_H	p_H	F	$F_{кр}$	S^2_H	p_H	F	$F_{кр}$
H_1	1470	7	4,4	2,1	406	11	28,2	2,7
H_2	960	20	3,2	3,3	253	16	37,3	3,6
H_3	—	—	—	—	119	15	4,75	3,3

Рассмотрим сначала задачу медицинского прогнозирования. Из таблицы следует, что гипотезу H_1 о линейной связи следует отклонить, так как величина F в два с лишним раза превосходит значение $F_{кр}$. Для гипотезы H_2 значение F не превышает значения $F_{кр}$, т. е. гипотеза не отклоняется. Другими словами, введение в формулу прогноза функции z вряд ли целесообразно.

Проверка гипотезы H_3 не имеет смысла, так как при отсутствии в (4) функции z гипотеза H_3 совпадает с гипотезой H_1, которую мы уже отклонили. Таким образом, можно заключить, что в данной задаче для прогноза тяжести состояния больного целесообразно использовать выражение (3).

Рассмотрим теперь задачу сейсмического прогнозирования. Из таблицы видно, что для каждой из гипотез значение F превышает значение $F_{кр}$, поэтому все они отклоняются. Следовательно, представляется оправданным использование выражения (1) для прогнозирования энергий землетрясений на Кавказе.

ЛИТЕРАТУРА

1. Сб. «Нелинейные и линейные методы в распознавании образов». М., «Наука», 1975.
2. *Р. С. Виницкая, И. Т. Турбович, А. А. Адамян, Е. Ф. Юрков, М. Я. Авруцкий, Ю. Д. Кашин.* Прогноз дыхательной недостаточности после операции на легких по данным дооперационных исследований.— Экспериментальная хирургия и анестезиология, 1973, № 5.
3. *В. И. Бунэ, И. Т. Турбович, Б. А. Борисов, В. Г. Гитис, Г. И. Рейснер, Е. Ф. Юрков.* О методе выявления связи максимально возможной магнитуды землетрясений с тектоническими параметрами района.— ДАН СССР, 1974, т. 214, № 3.
4. *М. Кендалл, А. Стьюарт.* Статистические выводы и связи. М., «Наука», 1973.
5. *Б. Л. ван дер Варден.* Математическая статистика. М., ИЛ, 1960.
6. *G. Box.* Fitting empirical data.— Ann. N. Y. Acad. of Sci., 1960, v. 86, art.3.
7. *Д. Худсон.* Статистика для физиков. М., «Мир», 1970.

И. Т. Турбович, Р. С. Виницкая, В. Г. Гитис, С. Г. Ерамян, В. С. Нагорнов, Н. Н. Сунгурян, Е. Ф. Юрков

Анализ связи между тяжестью состояния больного и его физиологическими показателями

В практике лечения ряда заболеваний часто возникает необходимость выяснить связь между тяжестью состояния больного и его физиологическими показателями. Знание такой связи могло бы быть полезным как для врача-клинициста, обладающего определенным опытом оценки состояния больного, так и для физиолога, занимающегося выяснением информативности измеряемых показателей и желающего минимизировать их число при разного рода экспертизах.

Качественный характер связи между клиническим состоянием больного и физиологическими показателями в общем известен: чем тяжелее состояние больного, тем в большей степени нарушены его физиологические показатели. Представляет интерес разработать подход к нахождению количественной связи между общей клинической оценкой состояния больного и объективно измеряемыми физиологическими показателями.

Попытка разработки такого подхода излагается в настоящей статье на примере задачи нахождения связи между тяжестью клинического состояния и показателями внешнего дыхания у больных хронической пневмонией.

В процессе исследования решались следующие вопросы: 1) выбор математической модели связи, 2) отбор наиболее существенных физиологических показателей, 3) проверка возможности упрощения выбранной модели связи.

Экспериментальный материал был взят из историй болезней 145 больных, обследованных и лечившихся в клинической больнице Государственного института усовершенствования врачей г. Еревана. По тяжести состояния больные были разделены на 5 групп на основании клинических признаков нарастающей легочной недостаточности [1].

В качестве объективно измеряемых физиологических показателей, отражающих нарушения внешнего дыхания, взяты: 1) число дыханий в минуту, 2) минутный объем дыхания (МОД), 3) коэффициент использования кислорода ($КИО_2$), 4) прирост углекислоты во время выдоха в альвеолярном газе ($\Delta CO_{2\Delta}$), 5) жизненная емкость легких (ЖЕЛ), 6) отношение резервных объемов выдоха и вдоха ($РО_{выд}/РО_{вд}$), 7) объем форсированной жизненной емкости легких за первую сек ($ФЖЕЛ_1$), 8) максимальная вентиляция легких (МВЛ), 9) мощность вдоха ($М_{вд}$), 10) мощность выдоха ($М_{выд}$). Все показатели за исключением $ФЖЕЛ_1$, $\Delta CO_{2\Delta}$ и $КИО_2$ были взяты в процентах к норме [1].

Выбор математической модели и нахождение связи. В настоящее время не существует общепринятой модели, достаточно точно отражающей характер связи физиологических показателей с клинической оценкой тяжести состояния больного. В этой ситуации при построении математической модели связи мы исходили из следующего принципа: выбрать наиболее простой класс функций, учитывающий все качественные соображения, известные врачу о характере этой связи, и затем ввести все обоснованные с медицинской точки зрения ограничения, накладываемые на параметры этих функций.

При выборе класса функций было использовано предположение о нелинейности связи. Врачебный опыт показывает, что равное изменение показателя в разных зонах его значений в общем случае приводит к неодинаковым изменениям тяжести состояния больного.

Достаточно простым классом функций, учитывающим допущение о нелинейности, является класс, представленный в виде суммы одномерных нелинейных функций [2]:

$$f(x_1, \ldots, x_I) = \sum_{i=1}^{I} \varphi_i(x_i),$$

где x_i — значение i-го показателя, $\varphi_i(x_i)$ — одномерная (нелинейная) функция, I — число рассматриваемых показателей.

При выборе ограничений на параметры одномерных нелинейных функций $\varphi_i(x_i)$ было использовано предположение о монотонности связи и о границах нормы показателей. Врачебный опыт показывает, что чем сильнее отклонен показатель от нормы, тем

---

[1] Нормальной (должной) величиной называется среднее значение показателя для здорового человека в соответствии с его ростом, полом, возрастом. Границами нормы для легочных объемов считают отклонения на $\pm 15\%$ ($\pm 20\%$) от средней должной величины, принимаемой за 100%.

при прочих равных условиях тяжелее состояние больного. Изменение же показателей в границах нормы практически не связано с изменением тяжести состояния больного. Поэтому допустимыми считались функции $\varphi_i(x_i)$, не уменьшающиеся при отклонении показателя от нормы и постоянные в границах нормы.

Модель связи между тяжестью состояния и физиологическими показателями, выбранная с учетом сформулированных предпосылок, имела следующий вид:

$$r_n = \sum_{i=1}^{I} \varphi_i(x_{ni}) + \varepsilon_n \quad (n = 1, \ldots, N),$$

где r_n — тяжесть (номер группы) состояния для n-го больного; x_{ni} — значение i-го показателя для n-го больного; ε_n — ошибка связи для n-го больного; N — общее число больных.

Функции $\varphi_i(x_i)$ вычислялись по выборке больных в классе кусочно-линейных функций по методу наименьших квадратов.

Одновременно с поиском функций производился отбор показателей x_i, наиболее существенно связанных с тяжестью состояния r. Для этой цели использован шаговый алгоритм выбора показателей [3]. При этом заключение о существенности выбранного набора показателей выносилось на основании F-критерия [3, 4].

При проверке возможности упрощения выбранной модели конкурирующей моделью, содержащей меньшее число подбираемых параметров, являлась модель, предполагавшая линейность функций $\varphi_i(x_i)$ за пределами нормы. Проверка этого предположения также производилась на основе F-критерия.

Как уже упоминалось ранее, в данной задаче тяжесть клинического состояния больного оценивалась по условной дискретной шкале: 1, 2, 3, 4, 5. Можно предполагать, что, получив оценку связи $f(x) = \sum_{i=1}^{I} \varphi_i(x_i)$, врач будет округлять ее до ближайшего целого с тем, чтобы полученная оценка также имела дискретный вид. В связи с этим меру качества прогноза было бы целесообразно связать не только с «теснотой» корреляции r и f, но и с числом ошибок, полученных при округлении значений $f(x)$ до ближайшего целого. Это обстоятельство учитывается в модифицированном методе наименьших квадратов [5]. Он позволяет за счет некоторого ухудшения абсолютной точности приближения (остаточной суммы квадратов) получить меньшее количество ошибок при округлении прогноза до ближайшего целого при неизменном коэффициенте корреляции r и f. Для этого функции $\varphi_i(x_i)$, получаемые по методу наименьших квадратов, достаточно откорректировать по весьма простым формулам.

Обсуждение результатов. На рис. 1 приведены графики зависимости F-отношения от числа показателей, отбираемых шаговым алгоритмом регрессионного анализа, и график критических значений $F_{кр}$ для 10%-го уровня значимости. Из рисунка видно, что

F-отношение становится ниже 10%-го уровня $F_{кр}$, начиная с трех отобранных показателей. Это означает, что гипотеза о существенности первых трех показателей, отобранных алгоритмом, может быть принята на 10%-ом уровне значимости.

Тремя показателями (в порядке важности, в котором они были отобраны алгоритмом) оказались: ФЖЕЛ$_1$, КИО$_2$ и М$_{выд}$. На

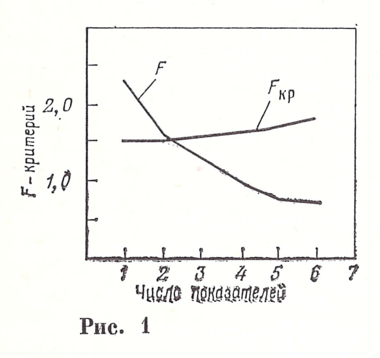


Рис. 1

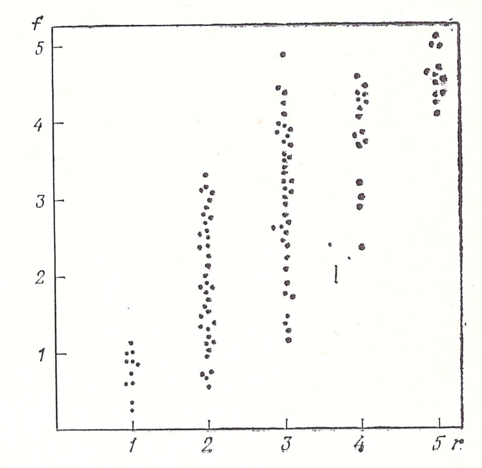


Рис. 2

Рис. 3

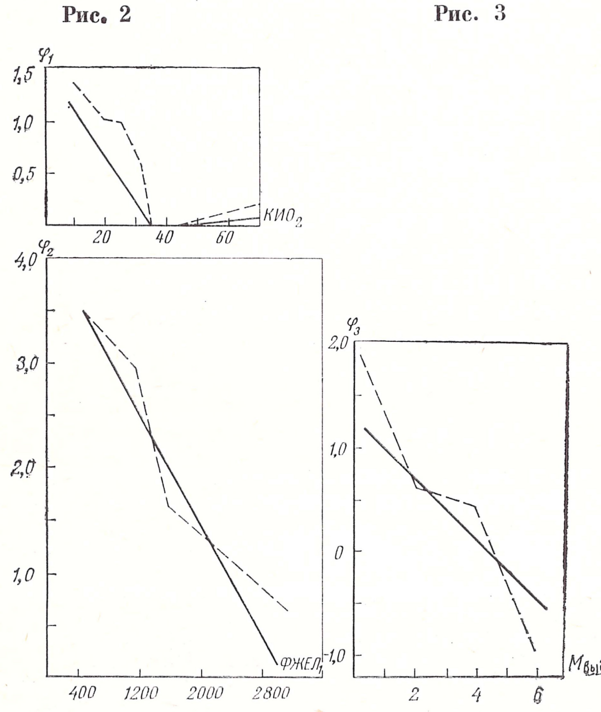


рис. 2 графики функций $\varphi_i(x_i)$ для этих показателей даны пунктиром. Отобранные по статистическому критерию физиологические показатели не противоречат медицинской точке зрения о важности этих показателей внешнего дыхания для оценки легочной недостаточности данного заболевания.

Для отобранных показателей на основании F-критерия была проверена гипотеза о линейности функций $\varphi_i(x_i)$ за пределами нормы. Значение F-отношения для этой гипотезы оказалось равным 1,14, что не превосходит 10%-й уровень $F_{кр}$, равный 1,5. Следовательно, гипотеза о линейности функций $\varphi_i(x_i)$ за пределами нормы не отвергается. Соответствующие этому случаю графики $\varphi_i(x_i)$ для трех указанных показателей изображены на рис. 2 непрерывными линиями. Из рис. 2 видно, что нелинейные функции весьма близки к линеаризованным функциям. В то же время вывод о том, что функции $\varphi_i(x_i)$ являются линейными за пределами нормы нельзя признать окончательным из-за малого экспериментального материала.

Сравнительная диаграмма оценок r и оценок f, полученных с помощью модифицированного метода наименьших квадратов, приведена на рис. 3. Оценки f рассчитаны для функций $\varphi_i(x_i)$, линейных за пределами нормы. Коэффициент корреляции между f и r оказался равным 0,82. Аналогичные результаты были получены при разбиении экспериментальной выборки на учебную и контрольную серию. По данным рис. 3 было подсчитано число больных, у которых округленная до ближайшего целого оценка f отличалась от клинической оценки r на единицу. Число таких больных оказалось равным 58 (44% от общего числа больных). Для трех больных расхождение между оценками было равно двум. Для остальных больных оценки совпадают.

Приведенные результаты показывают, что между клинической оценкой тяжести состояния и физиологическими показателями внешнего дыхания данной категории больных имеется достаточно тесная связь. Об этом говорит относительно высокое значение коэффициента корреляции, а также малое количество грубых ошибок (т. е. случаев, для которых округленное значение f отличается от клинической оценки r более чем на единицу).

ЛИТЕРАТУРА

1. *С. Г. Ерамян.* Вопросы легочно-сердечной недостаточности при различных хронических неспецифических заболеваниях легких. Автореф. докт. дис. Ереван, 1969.
2. Сб. «Нелинейные и линейные методы в распознавании образов». М., «Наука», 1975.
3. *Н. Дрейпер, Г. Смит.* Прикладной регрессионный анализ. М., «Статистика», 1973.
4. *Е. Ф. Юрков.* Статистический анализ модели прогнозирования, использующей одномерные нелинейные преобразования.— Наст. сб.
5. *Е. Ф. Юрков, В. С. Нагорнов.* Модифицированный метод наименьших квадратов при решении задач прогноза с дискретной прогнозируемой величиной.— Наст. сб.

Т. А. Бадалов

Применение метода одномерных нелинейных преобразований для прогнозирования естественного режима нефтяного пласта

Среди важнейших факторов, связанных с эксплуатацией нефтяных скважин, следует отметить режим пласта. Режим нефтяного пласта характеризует вид пластовой энергии, которая обусловливает приток жидкостей и газа к эксплуатирующимся скважинам. Он определяется как физико-геологическими природными условиями, так и условиями разработки и эксплуатации месторождения [1].

В связи с составлением технологической схемы весьма важно на ранней стадии разработки определить характер источников пластовой энергии, которыми располагает месторождение (естественный режим пласта). Определение естественного режима пласта сильно затруднено из-за отсутствия в этот период промысловых данных, характеризующих изменение пластового давления, дебита нефти, газа, воды, динамику продвижения контура нефтеносности и т. д. В начальной стадии разработки о возможном режиме пласта можно судить только по геолого-физическим свойствам пород-коллекторов и свойствам насыщающих флюидов.

Настоящая работа проведена с целью оценки возможности прогнозирования естественного режима пласта по совокупности геолого-физических признаков, имеющихся в начальный период разработки нефтяных месторождений, а также для установления связи между прогнозируемой величиной и признаками.

В качестве экспериментального материала были использованы данные Института проблем глубинных нефтегазовых месторождений АН Азербайджанской ССР о нефтяных залежах Азербайджана, находящихся в завершающей стадии эксплуатации. Задача нахождения связи между естественным режимом пласта и геолого-физическими признаками решалась с помощью метода одномерных нелинейных преобразований, разработанного в Институте проблем передачи информации АН СССР [2—6].

В процессе решения задачи анализировалось 90 залежей нефти основных месторождений Апшеронского полуострова: Балаханы-Сабунчи-Раманинского, Сураханского, Биби-Эйбатского, Калинского, Бинагадинского, Маштаги-Бузовнинского, Чахнаглярского, Сулутепинского, Шабандагского. Разработка этих залежей осуществлялась при очень уплотненной сетке скважин (1—2 га/скважину) на естественном режиме.

Исследовалось влияние на режим основных геолого-физических свойств пород-коллекторов и пластовой нефти, таких, как: 1) вязкость нефти; 2) логарифм проницаемости пород-коллекторов; 3) расчлененность разреза; 4) песчанистость в процентах;

5) цементация (IV фракция гранулометрического анализа + карбонатность) в процентах.

Для прогнозирования использовалась следующая классификация естественных режимов нефтяных залежей: 1) водонапорный режим (I класс); 2) смешанный режим (II класс); 3) режим растворенного газа (III класс).

При этом прогнозируемое состояние (режим) связывалось с числовой величиной r — номером класса. Исходя из общих представлений о влиянии основных геолого-физических свойств пород-коллекторов и пластовой нефти на естественный режим залежи, можно предположить, что при такой последовательности расположения классов их номера монотонно зависят от геолого-физических признаков. Так, например, из [1] следует, что уменьшение песчанистости при прочих одинаковых условиях связано с увеличением номера класса.

При прогнозировании предполагалось, что связь между естественным режимом пласта и геолого-физическими признаками описывается следующей моделью:

$$f(x_1, x_2, \ldots, x_I) = \sum_{i=1}^{I} \varphi_i(x_i),$$

где $\{x_i\}$ — геолого-физические признаки пласта; φ_i — искомые нелинейные одномерные монотонные функции; i — номер признака.

Функции φ_i находятся из условия минимума математического ожидания квадрата разности между прогнозируемой r и прогнозирующей f величинами. Алгоритм нахождения функций φ_i изложен в [3].

Руководствуясь соображениями выбора более простой модели, рассматривалась эффективность линейной модели по сравнению с указанной моделью. Такое сравнение было произведено на основе использования F-критерия [6, 7]. Оказалось, что дисперсионное отношение F, равное 4,86, превышает значение $F_{кр} = 1{,}98$ на уровне значимости 5%. На этом основании гипотеза о линейном характере зависимости была отвергнута.

При решении данной задачи наряду с получением прогноза было важно определить наиболее существенные геолого-физические признаки. С этой целью применялся алгоритм шагового выбора параметров [7], в котором оценка значимости признаков производилась по F-критерию. По степени важности признаки оказались расположенными в следующем порядке (в скобках указано дисперсионное отношение F): песчанистость (75,4), вязкость (23,5), расчлененность (1,9), проницаемость (2,1), цементация (1,7). При включении в модель трех последних признаков дисперсионное отношение F становится меньше $F_{кр} \approx 2{,}73$, что говорит о несущественности этих признаков для прогнозирования. Таким

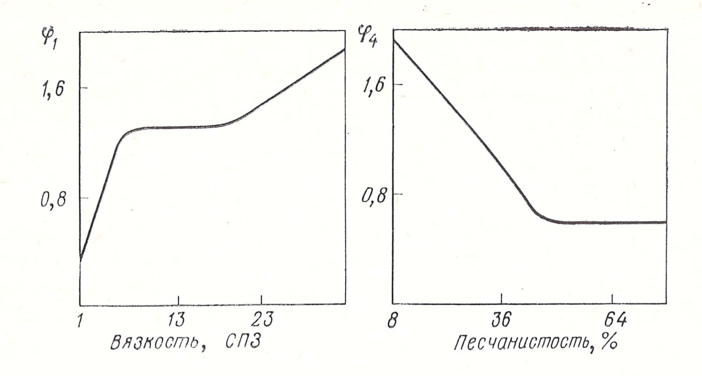


Рис. 1

образом, для прогнозирования достаточно взять два признака: вязкость и песчанистость.

Графики полученных функций φ_i, соответствующих признакам вязкости и песчанистости, приведены на рис. 1 (номера функций на рисунке соответствуют первоначальным номерам признаков). Эти графики имеют существенно нелинейный характер.

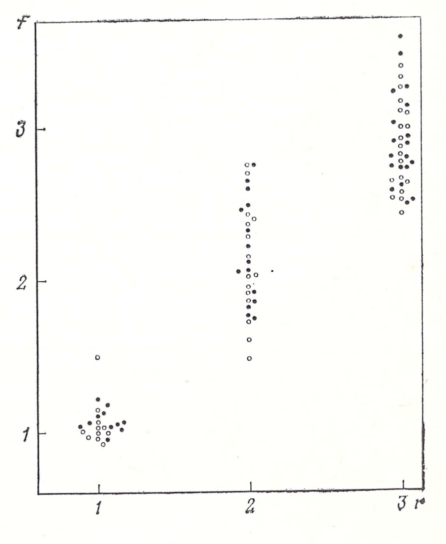


Рис. 2

График функции φ_1, связанный с вязкостью, условно можно разбить на три участка: средний — в пределах которого изменение вязкости не приводит к изменению прогноза, правый и левый — где с увеличением вязкости влияние функции φ_1 на прогноз возрастает.

Уменьшение песчанистости до 50%, как видно из графика φ_4 на рис. 1, не вызывает изменения в режиме нефтяных пластов, однако при дальнейшем снижении влияние песчанистости на прогноз проявляется в весьма сильной степени.

Качество прогнозирования оценивалось нормированной среднеквадратичной ошибкой прогнозирования

$$R = \left(\sum_{i=1}^{N} (r_i - f_i)^2 \Bigg/ \sum_{i=1}^{N} (r_i - \bar{r})^2 \right)^{1/2},$$

где r_i и f_i— экспериментальное и предсказанное значение прогно-

зируемой величины для i-й реализации выборки; $\bar{\bar{r}}$ — среднее экспериментальное значение прогнозируемой величины; N — число реализаций в экспериментальной выборке.

Значение R оказалось равным 0,386. Для дополнительной проверки качества прогнозирования экспериментальный материал был разбит на учебную и экзаменационную выборки и на основе прогнозирующих функций φ_1 и φ_4, вычисленных на материале учебной выборки, был осуществлен прогноз для этих выборок. Далее результат прогнозирования сопоставлялся с режимами, фактически существующими в пластах. Результаты такого сравнения приведены на рис. 2, где кружочками обозначены реализации учебной выборки, точками — экзаменационной. Из рисунка видно, что естественный режим нефтяного пласта с достаточной точностью можно прогнозировать по данным геолого-физического описания, имеющимся в начальный период разработки.

ЛИТЕРАТУРА

1. *М. А. Жданов, А. А. Карцев.* Нефтепромысловая геология и гидрогеология. М., Гостоптехиздат, 1958.
2. *И. Т. Турбович.* О нахождении скрытых закономерностей на основе опытных данных.— Сб. «Нелинейные и линейные методы в распознавании образов». М., «Наука», 1975.
3. *Е. Ф. Юрков.* Нахождение одномерных нелинейных преобразований на основе одномерных статистических характеристик при прогнозировании.— Сб. «Нелинейные и линейные методы в распознавании образов». «Наука», 1975.
4. *В. Г. Гитис.* Алгоритмы прогнозирования и синтеза признаков с использованием одномерных кусочно-линейных функций.—Сб. «Нелинейные и линейные методы в распознавании образов». М., «Наука», 1975.
5. *В. Г. Гитис, В. С. Нагорнов, И. Т. Турбович, Е. Ф. Юрков.* Решение задач медицинского и сейсмического прогнозирования на основе использования одномерных нелинейных преобразований.— Сб. «Нелинейные и линейные методы в распознавании образов». М., «Наука», 1975.
6. *Е. Ф. Юрков.* Статистический анализ модели прогнозирования, использующей одномерные нелинейные преобразования.— Наст. сб.
7. *Н. Дрейпер, Г. Смит.* Прикладной регрессионный анализ. М., «Статистика», 1973.

А. П. Вайншток, Е. И. Кабанова, В. А. Махонин

К использованию алгебраических приемов идентификации систем для анализа речи

Для улучшения качества речи, искаженной помехами, и для представления элементов слитной речи нужно идентифицировать акустическую модель речеобразования и входные сигналы этой модели. Нестационарный и нелинейный характер точной модели вынуждает исследователей искать такие представления речевого

сигнала, которые допускали бы артикуляторную интерпретацию и простые процедуры оценивания параметров приближенных моделей. Способ построения упрощенной модели, несмотря на кажущуюся доступность анализу артикуляторных действий говорящего, до сих пор остается предметом споров. Единственное, что, как кажется, начинают признавать все исследователи,— это неадекватность энергетического спектрополосного препарирования тонкой временной структуры речевого сигнала. Предположение о существенности медленных изменений параметров колебательных компонент и несущественности тонкой временно́й структуры, как не влияющей на разборчивость речи, передаваемой по телефонному каналу, оказалось безвредным лишь в технике синтетической телефонии. Многие из описанных явлений, сопровождающих изменения темпа речи, оказались артефактами сонографа.

Попытки автоматического выделения сочетаний сигналов, отождествляемых с признаками фонем, продолжаются, но уже без особого энтузиазма. Хорошо воспроизводимыми элементами слитной речи оказались фрагменты сигнала, близкие к слогам, особенно типа согласный — гласный (СГ). Сообщения об успешных опытах по автоматическому обнаружению этих слогов поступают из многих лабораторий [1—3].

Стремление повысить разрешающую способность анализаторов состава сигнала привело к переходу от классических методов анализа непрерывных сигналов к современным приемам дискретного анализа, в частности к алгебраическому аппарату теории систем [4, 8]. Использование рекуррентных соотношений этой теории позволяет ограничить время формирования суммы e колебаний до $k = 2e + 1$ отсчетов дискретной шкалы времени. Этот предел быстродействия достигается распараллеливанием обработки, несколько отсчетов сигнала выступают при этом как единый вектор. Характеристики системы представляют при этом переходной матрицей, формула преобразования эквивалентна предсказанию

$$\left\| \begin{matrix} u_{n-k+1} \\ u_{n-k+2} \\ \vdots \\ u_{n-1} \\ u_n \end{matrix} \right\| = \left\| \begin{matrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ a_p & \dots & & & a_2 a_1 \end{matrix} \right\| \left\| \begin{matrix} u_{n-k} \\ u_{n-k+1} \\ \vdots \\ u_{n-2} \\ u_{n-1} \end{matrix} \right\|. \qquad (1)$$

Предсказание $u_n = \sum_{i=1}^{p} a_i u_{n-i}$ требует знания вектора коэффициентов — $\boldsymbol{a}$. Проверку адекватности представления модели порождения сигнала линейной системой со стационарной переходной матрицей можно произвести без вычисления коэффициентов предсказания. Достаточно заполнить отсчетами сигнала матрицу Ган-

келя и вычислить ее определитель

$$\begin{vmatrix} u_n u_{n-1} \ldots u_{n-p-1} u_{n-p} \\ u_{n-1} u_{n-2} \ldots u_{n-p-2} u_{n-p-1} \\ \ldots\ldots\ldots\ldots\ldots \\ u_{n-p} u_{n-p-1} \ldots u_{n-2p-1} u_{n-2p} \end{vmatrix} \quad (2)$$

Если определитель этой матрицы равен нулю, то это означает, что последние $2p+1$ отсчетов сигнала порождались линейной системой со стационарной переходной матрицей. Практически значения сигнала искажаются помехой, что приводит к флуктуациям значения этого определителя даже там, где порядок системы достаточен для представления текущих отрезков сигнала. В те моменты времени, когда условие достаточности не выполняется, этот определитель уклоняется от нуля, образуя характерные всплески. Самые мощные всплески отмечают моменты смыкания голосовых связок, где резко меняются как начальный вектор, так и порядок системы. Моменты размыкания связок, подключения и отключения носовой полости сопровождаются менее интенсивными всплесками.

Так же, как и в исследованиях подобного метода, основанного на вычислении определителя Грама [5, 6], возникают трудности с определением допуска на уклонение определителя от нуля, так как динамический диапазон значений определителя очень велик. Были предприняты вычислительные эксперименты с использованием перемешивания отсчетов [7] перед заполнением определителя и использованием определителя с перемешанными отсчетами при задании доверительных интервалов для значений неперемешанного определителя.

Если определитель (2) разложить по элементам первой строки (или столбца), то получится уравнение линейного предсказания

$$\tilde{u}_n A_{11} + u_{n-1} A_{12} + \ldots + u_{n-p} A_{1p} = 0,$$

где A — адъюнкты.

Сравнение u_n и $\tilde{u}_n$ и нескольких других комбинаций, предсказанных $\tilde{u}_{n+1}$, $\tilde{u}_{n+2}$, с истинными позволяет вычленить в сигнале моменты смыкания голосовых связок и отрезки свободных колебаний значительно увереннее, чем непосредственно по полному определителю ганкелевой матрицы, и с большей точностью, чем при использовании определителя Грама. Пример членения приведен на графиках рис. 1, где показаны осциллограммы сигнала, кривые определителей ганкелевой матрицы разных порядков, что нужно для оценки ранга переходной матрицы порождающей системы, и ошибки предсказания. Эти кривые указывают на полезность кусочно-стационарного представления акустической модели речи, параметры которой должны оцениваться синхронно с работой голосовых связок.

Артикуляторные движения, пренебрегая изменениями на участке свободных колебаний, можно описать моделью, определяю-

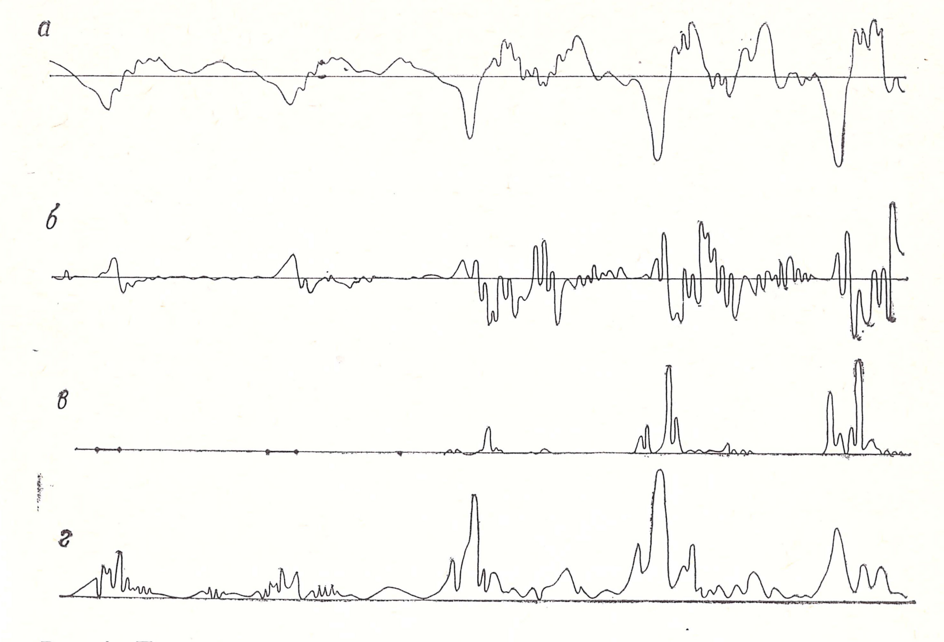


Рис. 1. Пример членения речевого сигнала

а — участок перехода в сочетании *ну*; *б* — значение определителя ганкелевой матрицы 3-го порядка; *в* — модуль значения определителя ганкелевой матрицы 7-го порядка; *г* — модуль ошибки предсказания

щей эти движения, можно продолжить это расчленение моделей и далее, выделяя блоки, описывающие отдельные функции, и блоки, обеспечивающие взаимодействие этих функций. При такой декомпозиции всегда сохранится блок, имитирующий акустические процессы в речевом тракте. Для ускорения самонастройки этой модели в свою очередь потребуются измерительные процедуры, способные оценивать отдельные параметры формантных колебаний независимо от всех остальных и на коротком отрезке сигнала. Геометрический подход здесь также оказывается эффективным.

Рассмотрим, например, одну колебательную компоненту $x(n\tau) = Ae^{-\delta n\tau}\cos(\omega n\tau + \psi)$. Значения ее отсчетов связаны с принятыми параметрами колебаний ω, δ следующими соотношениями:

$$x_n = a_1 x_{n-1} + a_2 x_{n-2}, \qquad a_1 = 2e^{-\delta\tau}\cos\omega\tau, \qquad a_2 = -e^{-2\delta\tau},$$

$$A\xi_{n-1} = \begin{Vmatrix} 0 & 1 \\ a_2 & a_1 \end{Vmatrix} \begin{Vmatrix} x_{n-1} \\ x_n \end{Vmatrix} = \begin{Vmatrix} x_n \\ x_{n+1} \end{Vmatrix} = \xi_n.$$

По записи колебания $x(n\tau)$ можно сформировать последовательность векторов $\xi_1, \xi_2, \ldots, \xi_n$. Тогда на плоскости с координатами (x_n, x_{n+1}) синусоидальное колебание изобразится эллипсом, а экспоненциально модулированное синусоидальное колебание —

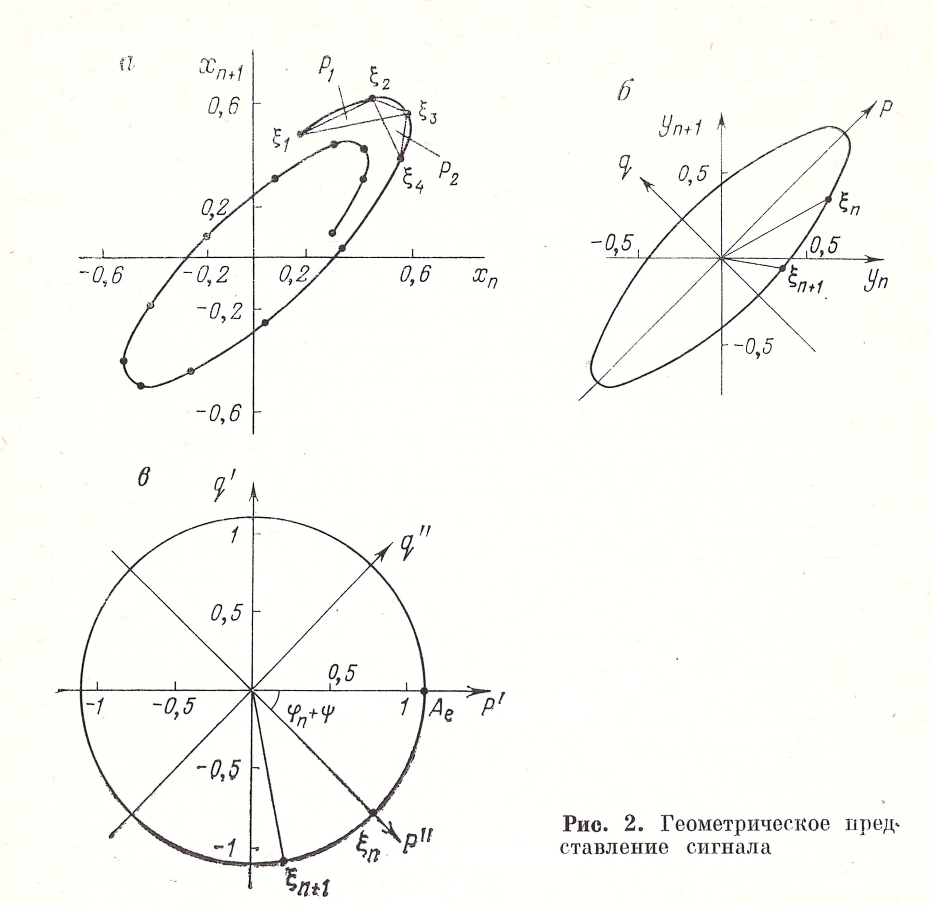


Рис. 2. Геометрическое представление сигнала

спиралью. Геометрические характеристики этих кривых можно использовать для оценивания параметров по короткому отрезку сигнала.

1. Оценка постоянной времени δ. Площадь треугольника, образованного на плоскости $(x_n,\ x_{n+1})$ концами векторов ξ_n, ξ_{n+1}, ξ_{n+2} (рис. 2, *а*), записывается через отсчеты сигнала $x\,(n\tau)$ по формуле

$$P_n = \begin{vmatrix} x_n & x_{n+1} & 1 \\ x_{n+1} & x_{n+2} & 1 \\ x_{n+2} & x_{n+3} & 1 \end{vmatrix}.$$

Отношение площадей P_n и P_{n+1} на спирали зависит от постоянной времени δ: $P_{n+1}/P_n = e^{-2\delta\tau}$.

Для повышения помехозащищенности на отрезке из N отсчетов сигнала постоянная времени определяется с накоплением до

деления по формуле

$$e^{-2\delta\tau} = \sum_{n=1}^{N-5} P_{n+1}P_{n+2} \Big/ \sum_{n=1}^{N-5} P_n P_{n+2}.$$

Полученную оценку можно использовать для компенсации затухания. Остальные параметры определяются по незатухающему колебанию $y(n\tau) = x(n\tau)e^{\delta n\tau} = A\cos(\omega n\tau + \psi)$.

2. Оценка амплитуды A. Для последовательных отсчетов синусоидального колебания $y(n\tau)$ справедливы выражения $y_{n+1}^2 - y_n y_{n+2} = A^2\sin^2\omega\tau$, $y_n y_{n+3} - y_{n+1}y_{n+2} = -2A^2\sin^2\omega\tau\cos\omega\tau$, откуда

$$(y_n y_{n+3} - y_{n+1}y_{n+2})/(y_{n+1}^2 - y_n y_{n+2}) = -2\cos\omega\tau.$$

Мгновенное значение амплитуды определяется из выражения

$$A^2 = \frac{4(y_{n+1}^2 - y_n y_{n+2})^3}{4(y_{n+1}^2 - y_n y_{n+2})^2 - (y_n y_{n+3} - y_{n+1}y_{n+2})^2} = \frac{U_n}{V_n}.$$

На отрезке из N отсчетов амплитуда оценивается по формуле

$$A^2 = \sum_{n=1}^{M} U_n V_{n+1} \Big/ \sum_{n=1}^{M} V_n V_{n+1}, \qquad (3)$$

где $M = N - 3$.

3. Оценка частоты ω. На плоскости (y_n, y_{n+1}) синусоидальное колебание $y(n\tau)$ изображается эллипсом (рис. 2, *б*), большая ось которого повернута относительно оси y_n на угол $\pi/4$. Соотношение полуосей эллипса определяется частотой ω и интервалом дискретизации τ. Перейдем к системе координат (p, q): $p_n = y_n/\sqrt{2} + y_{n+1}/\sqrt{2}$, $q_n = -y_n/\sqrt{2} + y_{n+1}/\sqrt{2}$. По двум точкам эллипса $\xi_n(p_n, q_n)$ и $\xi_{n+1}(p_{n+1}, q_{n+1})$ можно определить отношение его полуосей $(B_e/A_e)^2 = (q_{n+1}^2 - q_n^2)/(p_n^2 - p_{n+1}^2)$. Подставив в эту формулу выражения для координат p и q, получим

$$\operatorname{tg}^2(\omega\tau/2) = (-y_n + 2y_{n+1} - y_{n+2})/(y_n + 2y_{n+1} + y_{n+2}) = U_n/V_n.$$

Для N отсчетов сигнала величина $\operatorname{tg}^2(\omega\tau/2)$ определяется по выражению, аналогичному (3) при $M = N - 2$.

4. Оценка начальной фазы ψ. От последовательности векторов ξ_n с координатами (p, q) можно перейти к последовательности векторов ξ_n' на плоскости (p', q') (рис. 2, *в*), при этом эллипс преобразуется в окружность. Координаты векторов связаны соотношениями $p_n' = p_n$, $q_n' = q_n/\operatorname{tg}(\omega\tau/2)$.

Если затем сделать поворот осей координат на угол $-(\varphi_n + \psi) = -[\omega\tau/2 + (n-1)\omega\tau + \psi]$ и перейти к координатам (p'', q''), то получим для вектора ξ_n'' координаты $p_n'' = A_e$, $q_n'' = 0$ или

в развернутом виде

$$p_n' \cos(\varphi_n + \psi) - q_n' \sin(\varphi_n + \psi) = A_e,$$

$$p_n' \sin(\varphi_n + \psi) + q_n' \cos(\varphi_n + \psi) = 0.$$

Записав аналогичную систему уравнений для вектора ξ_{n+1}'' и подставив выражения для p_n' и q_n' через отсчеты сигнала $y(n\tau)$, получим

$$\cos(\varphi_n + \psi) = \frac{\sqrt{2}\, A_e (y_n + y_{n+1}) \operatorname{tg}^2(\omega\tau/2)}{\operatorname{tg}^2(\omega\tau/2)(y_n + y_{n+1})^2 + (y_n - y_{n+1})^2} = P,$$

$$\cos(\varphi_{n-1} + \psi) = \frac{\sqrt{2}\, A_e (y_{n+1} + y_{n+2}) \operatorname{tg}^2(\omega\tau/2)}{\operatorname{tg}^2(\omega\tau/2)(y_{n+1} + y_{n+2})^2 + (y_{n+1} - y_{n+2})^2} = Q.$$

Используя формулу для косинуса суммы двух углов, можно получить мгновенную оценку начальной фазы ψ:

$$\operatorname{tg}\psi = \frac{P\cos\varphi_{n+1} - Q\cos\varphi_n}{P\sin\varphi_{n+1} - Q\sin\varphi_n} = U_n/V_n.$$

Для N отсчетов сигнала $\operatorname{tg}\psi$ определяется по формуле, аналогичной (3) при $M = N - 2$.

Процедуры измерения параметров компонент сигнала объединяются вместе с процедурами моделирования процесса генерирования сигнала, с процедурами обнаружения моментов смыкания голосовых связок и фильтрации компонент в схеме анализа через синтез, которую, как ожидается, можно будет дополнить моделями артикуляции, просодики и т. д. и получить универсальный алгоритм автоматического представления речевой деятельности человека.

Относительно структуры этого возможного алгоритма можно высказать следующие предположения. Поскольку произносительной единицей является слог, то объектами анализа должны быть план произнесения слога и те отдельные параметры, которые связаны с диалектными и с индивидуальными особенностями говорящего, с влиянием на слог более крупных произносительных единиц, таких, как слово, словосочетания и фразы.

Если все, что было связано с выбором слога и его параметров, рассматривать как постоянный вектор-параметр, задающий весь процесс на протяжении длительности слога (длительность входит как одна из величин, составляющих этот вектор-параметр), то отсюда последует невозможность линейного совместного представления и оценивания параметров и состояния [9]. Это означает, что все подходы к такой задаче должны быть типа настраиваемой модели, т. е. иметь итеративный характер. Схему настраиваемой модели можно составить так, чтобы отделить общечеловеческое от индивидуального. Так все фильтры, преобразователи и т. д. можно расчленить на пары блоков, один из которых — основной, опорный, «смысловой», другой — «индивидуальный», который

связывает конкретную реализацию сигнала с опорным блоком. Одно из требований к опорному блоку — удобство оценивания.

Такое представление парами блоков можно развить на всех уровнях начиная с акустики. Например, известны пересчеты параметров речевого тракта мужчины в параметры речевого тракта женщины или ребенка [10], коартикуляции, интонаций и т. д.

Реализация настраиваемой модели потребует создания эффективных вычислительных методов, в которых должна учитываться специфика речеобразования.

ЛИТЕРАТУРА

1. *А. В. Книппер, В. А. Махонин.* К описанию речевых сигналов.— Сб. «Речевое общение в автоматизированных системах». М., «Наука», 1975.
2. *S. Riviera.* Real-time automatic detection of syllable nuclei in continuous speech.— Speech communication Seminar, Stockholm, Aug. 1—3, 1974.
3. *O. Fujimura.* Syllable as unit of speech recognition.— IEEE Trans. Acoust. Speech and Signal Processing, v. ASSP-23 Febr. 1975.
4. *B. W. Dickinson, T. Kailath, M. Morf.* Canonical matrix function and state-space descriptions for deterministic and stockastic linear systems.— IEEE Trans., v. AC-19, N 6, 1974.
5. *А. Н. Собакин.* Об определении формантных параметров голосового тракта по речевому сигналу с помощью ЭВМ.— Акуст. журн., 1972, т. 18, № 1.
6. *H. W. Strube.* Determination of the instant of glottal closure from the speech wave.— JASA, 1974, v. 56, N 5.
7. *И. Ш. Пинскер.* Оценка метода обучения и обучающей выборки.— Сб. «Моделирование и автоматический анализ электрокардиограмм». М., «Наука», 1975.
8. *Б. И. Арнольд.* Обыкновенные дифференциальные уравнения. М., «Наука», 1971.
9. *П. Эйкхофф.* Основы идентификации систем управления. М., «Мир», 1975.
10. Quart. Progr. Status Report Speech Transmiss.— Lab. STL QPSR 2-3, 1975, Stockholm.

М. В. Назаров, Ю. Н. Прохоров

Алгоритм совместного рекуррентного оценивания параметров речевых сигналов

Разработка оптимальных методов сокращенного представления речевых сигналов может быть сведена к задаче синтеза устройств оптимальной идентификации исходной параметрической модели речеобразования. Наиболее эффективным методом при этом является одновременное совместное определение всех параметров модели по наблюдаемому сигналу. Вместе с тем в целях упрощения структуры синтезируемых устройств процедуру оценивания обычно разбивают на два этапа: идентификация модели речевого

тракта и оценивание параметров модели источника звукового возбуждения. Решение задач каждого этапа может производиться последовательно, когда на втором этапе используются результаты первого, и параллельно (независимо), когда задачи двух этапов рассматриваются как раздельные.

Естественно, что упрощение синтезируемых устройств при последовательном и параллельном двухэтапном оценивании достигается за счет некоторого снижения эффективности.

Процедуры оптимального независимого оценивания параметров рассматривались в литературе. Так, в работах [1, 3] сравнительно подробно исследованы вопросы синтеза оптимальных и субоптимальных устройств выделения основного тона, а также построения устройств выделения сигнала возбуждения следящего типа [2].

В ряде работ синтезированы схемы оптимального оценивания локально-постоянных параметров модели речевого тракта [4, 5].

Синтез алгоритма совместного оценивания в литературе не рассматривался. Вместе с тем такая задача представляет определенный интерес, так как ее решение устраняет известное противоречие между стремлением повысить точность оценивания параметров речевого тракта и необходимостью знания сигнала возбуждения для осуществления такого улучшения.

Для облегчения решения указанной задачи целесообразно использовать рекуррентные алгоритмы, которые удобно синтезировать, применяя теорию марковской фильтрации. Разработка оптимального устройства оценивания позволяет не только представить структуру оптимального анализатора, но и выявить наиболее целесообразные для практики упрощения. С таких позиций, например, были синтезированы упрощенные алгоритмы рекуррентного оценивания, рассмотренные в [6].

В одном из наиболее общих случаев речевой сигнал $x(t)$ и сигнал звукового возбуждения $f(t)$ связаны интегральным уравнением

$$x(t) = \int_0^t h_x(\vartheta_t, u)\, x(u)\, du + \int_0^t h_f(\vartheta_t, \tau)\, f(\tau)\, d\tau, \tag{1}$$

где $h_x(\vartheta_t, t')$, $h_f(\vartheta_t, t')$ — импульсные реакции цепи обратной связи и прямой цепи.

В несколько более простом варианте можно положить $h_f(\vartheta_t, t') = \delta(t' - t)$, тогда

$$x(t) = \int_0^t h_x(\vartheta_t, u)\, x(u)\, du + f(t). \tag{2}$$

Для параметризации уравнения (2) представим импульсную реакцию в виде ряда

$$h_x(\vartheta_t, t') = \vartheta_t^T \varphi(t'), \tag{3}$$

где $\vartheta_t = [\vartheta_1(t), \vartheta_2(t), \ldots, \vartheta_m(t)]^T$ — вектор параметров модели речевого тракта, $\varphi(t') = [\varphi_1(t'), \ldots, \varphi_m(t')]^T$, $\{\varphi_i(t)\}_{i=1}^m$ — ансамбль ортогональных с произвольным весом функций.

Модель (2) является довольно общей. Так, при выборе в качестве φ_i дельта-функций можно получить обычное уравнение авторегрессии — модель линейного предсказания. В случае, когда ортогональные функции обладают неперекрывающимися и примыкающими друг к другу спектрами, приходим к некоторому аналогу полосного вокодера [7].

Из (2), (3) следует

$$x_t = \vartheta_t^T z(x_0^t, t) + (1 - \nu) a_t s_t + b_t n_t, \tag{4}$$

где $x_t = x(t)$; x_0^t — траектория $x(t')$; $0 \leqslant t' \leqslant t$; $f_t = (1 - \nu) a_t s_t + b_t n_t = f(t)$ — сигнал звукового возбуждения с основным тоном s_t; $z(x_0^t, t) = [z_1, \ldots, z_m]^T$, $z_i(t) = \int_0^t \varphi_i(t - u) x(u) du$; a_t — процесс, управляющий амплитудой s_t; n_t — шумовое возбуждение — белый шум с дисперсией b_t^2;

$$\nu = \begin{cases} 1, & \text{если основной тон отсутствует.} \\ 0, & \text{если произносится озвученный участок речи.} \end{cases}$$

Наличие в модели (4) случайных процессов ϑ_t и n_t обусловливает стохастический характер x_t.

Наибольшие затруднения вызывает построение модели $a_t s_t$. Здесь возможны несколько подходов. Рассмотрим один из них. Представим сигнал основного тона в виде ряда Фурье

$$a_t s_t = \sum_{k=1}^{N} [A_k + m_k D_t^{(k)}] \sin(\omega_k t + \beta_t^{(k)} + \varphi_t^{(k)}), \tag{5}$$

где N — количество гармоник ряда, ω_k, A_k — частоты и амплитуды гармоник в отсутствии модуляции, $D_t^{(k)}$, $\beta_t^{(k)}$ — процессы, управляющие звуковым возбуждением при артикуляции, $\beta_t^{(k)} + \varphi_t^{(k)} = \Theta_t^{(k)}$, $\varphi_t^{(k)}$ — случайное колебание фазы гармоники за счет неточности модели, а также вследствие нестабильности опорных генераторов анализатора.

Введем следующие предположения о ϑ_t, $\beta_t^{(k)}$, $\varphi_t^{(k)}$, $D_t^{(k)}$.

I. Параметры являются независимыми случайными процессами.

II. Параметры модели удовлетворяют следующим стохастическим уравнениям:

$$d\vartheta_t = -F_\vartheta \vartheta_t dt + G_\vartheta d\eta_\vartheta(t),$$
$$d\Theta_t = -F_\beta \beta_t dt + G_\beta d\eta_\beta(t) + G_\varphi d\eta_\varphi(t),$$
$$d\beta_t = -F_\beta \beta_t dt + G_\beta d\eta_\beta(t), \qquad d\boldsymbol{D}_t = -F_D \boldsymbol{D}_t dt + G_D d\eta_D(t), \tag{6}$$
$$d\varphi_t = G_\varphi d\eta_\varphi(t),$$

где Θ_t, β_t, D_t, φ_t — векторы-столбцы $N \times 1$; F_ϑ, F_β, F_D, G_ϑ, G_β, G_D, G_φ—диагональные матрицы с элементами, известными априори; $\eta_\vartheta(t)$, $\eta_\beta(t)$, $\eta_D(t)$, $\eta_\varphi(t)$ — векторы-столбцы ($N \times 1$) независимых винеровских процессов.

Правомерность такого предположения относительно параметров уравнения (5) в какой-то мере подтверждается экспериментальными данными [8]. В связной речи, однако, все параметры могут изменяться скачкообразно. В этом случае применение (6) приведет к некоторому «размазыванию» скачков, тем не менее переход к моделям с непрерывно изменяющимися параметрами удобно производить начиная с простых случаев типа (6).

III. Будем считать, что эффективная полоса пропускания цепи обратной связи схемы слежения за k-й гармоникой основного тона такова, что в полосу захвата попадает лишь одна гармоника.

Принятые предположения дают возможность решить задачу синтеза методами теории нелинейной марковской фильтрации в гауссовском приближении. Следует, однако, заметить, что это предположение правомерно здесь лишь в режиме высокой апостериорной точности, так как в (5) присутствует нелинейность. Построение оптимальных структур в негауссовском случае, как правило, приводит к необходимости значительного увеличения ансамбля оцениваемых параметров. Вследствие этого представляется целесообразным проведение синтеза анализатора для установившегося режима; начальные условия в этом случае могут задаваться некоторым устройством «грубого измерения», функционирующим на начальном интервале времени. В качестве такого устройства может быть использован нелинейный фильтр, разработанный для оценивания локально-постоянных параметров, или любое другое устройство (например, для получения начальной оценки мгновенной частоты основного тона можно использовать один из известных методов выделения сигнала основного тона).

Используя в рамках принятых ограничений методику теории нелинейной фильтрации [9], можно получить следующие уравнения для оценок искомых параметров и семиинвариантов апостериорного распределения:

$$\dot{m}_{\vartheta t} = -F_\vartheta m_{\vartheta t} + \frac{2K}{(1-\nu)\,b^2+\nu}\,\varepsilon_t z(x_0^t, t),$$

$$\dot{K} = \frac{1}{2}G_\vartheta - 2F_\vartheta K - \frac{2}{(1-\nu)\,b^2+\nu}\,K z(x_0^t, t)\,z^T(x_0^t, t)\,K^T,$$

$$\dot{m}_{\beta t}^{(k)} = -F_\beta^{(k)} m_{\beta t}^{(k)} + \frac{2(1-\nu)\,K_{\beta\Theta}^{(k)}}{(1-\nu)\,b^2+\nu}\,[A_k + m_k \hat{D}_t^{(k)}]\,\varepsilon_t \times$$
$$\times \cos(\omega_k t + \hat{\Theta}_t^{(k)}), \qquad (7)$$

$$\dot{\hat{\Theta}}_t^{(k)} = -F_\beta^{(k)} m_{\beta t}^{(k)} + \frac{2(1-\nu)\,K_{\Theta\Theta}^{(k)}}{(1-\nu)\,b^2+\nu}\,[A_k + m_k \hat{D}_t^{(k)}]\,\varepsilon_t \cos(\omega_k t + \hat{\Theta}_t^{(k)}),$$

$$\dot{\hat{D}}_t^{(k)} = -F_D^{(k)}\hat{D}_t^{(k)} + \frac{2(1-\nu)K_{DD}^{(k)}}{(1-\nu)b^2+\nu} m_k\varepsilon_t \sin(\omega_k t + \hat{\Theta}_t^{(k)}),$$

$$\dot{K}_{\beta\beta}^{(k)} = \frac{1}{2}G_\beta^{(k)} - 2F_\beta^{(k)}K_{\beta\beta}^{(k)} - \frac{1-\nu}{(1-\nu)b^2+\nu}[A_k + m_k\hat{D}_t^{(k)}](K_{\beta\Theta}^{(k)})^2,$$

$$\dot{K}_{\Theta\Theta}^{(k)} = \frac{1}{2}(G_\beta^{(k)} + G_\varphi^{(k)}) - 2F_\beta^{(k)}K_{\beta\Theta}^{(k)} - \frac{1-\nu}{b^2}[A_k + m_k\hat{D}_t^{(k)}]^2(K_{\Theta\Theta}^{(k)})^2,$$

$$\dot{K}_{\beta\Theta}^{(k)} = \frac{1}{2}G_\beta^{(k)} - F_\beta^{(k)}K_{\beta\Theta}^{(k)} - F_\beta^{(k)}K_{\beta\beta}^{(k)} - \frac{1-\nu}{(1-\nu)b^2+\nu}[A_k + m_k\hat{D}_t^{(k)}]^2 \times K_{\beta\Theta}^{(k)}K_{\Theta\Theta}^{(k)}, \qquad (8)$$

$$\varepsilon_t = x_t - \boldsymbol{m}_{\vartheta t}^T\boldsymbol{z}(x_0^t, t) - (1-\nu)\sum_{k=1}^{N}[A_k + m_k\hat{D}_t^{(k)}]\sin(\omega_k t + \hat{\Theta}_t^{(k)}),$$

где $\boldsymbol{m}_{\vartheta t}$, $m_{\beta t}^{(k)}$, $\hat{\Theta}_t^{(k)}$, $\hat{D}_t^{(k)}$ — оценки параметров $\boldsymbol{\vartheta}_t$, $\beta_t^{(k)}$, $\Theta_t^{(k)}$, $D_t^{(k)}$ (индекс k относится к номеру гармоники); K — матрица семиинвариантов оценки $m_{\vartheta t}$; $K_{\beta\beta}^{(k)}$, $K_{\beta\Theta}^{(k)}$, $K_{\Theta\Theta}^{(k)}$, $K_{DD}^{(k)}$ — семиинварианты оценок параметров сигнала основного тона; ν — случайная величина, оценивание которой может быть произведено с помощью известных методов автоматического выделения озвученных участков речи [1] или посредством контроля процесса $\boldsymbol{D}_t$.

Уравнение для семиинварианта $K_{DD}^{(k)}$ не приводится, так как не зависит от наблюдений и оценок других параметров.

Первые два уравнения (7) моделируются устройством, блок-схема которого приведена в [7]. Отличие от рассматриваемого случая заключается в способе формирования ε_t — необходимо учесть наличие оценки сигнала основного тона $\hat{a}_t\hat{s}_t$, которую следует вычесть из уклонения $x_t - \boldsymbol{m}_t^T\boldsymbol{z}(x_0^t, t)$. Построение блок-схем устройств, моделирующих остальные уравнения, не вызывает затруднений.

Отметим, что уравнения, определяющие оценку a_ts_t, описывают систему ФАПЧ, близкую к ранее применяемой для выделения основного тона в [1].

Полученные результаты могут быть обобщены на случай неизвестных априори матриц F_ϑ, F_β, F_D в соответствии с квазиоптимальной адаптивной нелинейной фильтрацией [10].

Для упрощения процедуры анализа приведенный способ совместного оценивания можно использовать на одном или нескольких этапах последовательного оценивания параметров сигнала [6]. В этом случае размерность вектора ϑ_t и матрицы K уменьшается, и алгоритм оказывается более простым.

ЛИТЕРАТУРА

1. *А. А. Пирогов* и др. Вокодерная телефония. М., «Связь», 1974.
2. *С. П. Баронин*, *А. И. Куштуев*. О построении выделителей основного тона речи следящего типа.— 8-я Всесоюзная конференция по акустике (рефераты), т. 1. М., 1973.
3. *С. П. Баронин*. Статистические методы анализа речевых сигналов. — Электросвязь, 1966, № 5.
4. *F. Itakura*, *S. Saito*. A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies.— Trans. Inst. Elect. Com. Engr. Jap., 53—A, 1, 1970.
5. *Matsui*, *Nakajima*, *Suzuki*, *Omura*. Speech Analysis by Kalman Filter Theory.— Bull. Electrotechn. Lab., 1972, v. 36, N 3.
6. *Ю. Н. Прохоров*. Рекуррентное оценивание параметров речевых сигналов.— Наст. сб.
7. *М. В. Назаров*, *Ю. Н. Прохоров*. Оптимальное оценивание параметров речевых сигналов.— Электросвязь, 1975, № 10.
8. *H. L. Schaffer*, *L. Dolansky*. On the Information Rate of Pitch Signals.— IEEE Int. Conv. Rec., pt 6, 1964.
9. *В. И. Тихонов*, *Ю. В. Саютин*. Синтез автоматических систем передачи информации с нелинейными фильтрами.— Сб. «Статистические методы в проектировании нелинейных систем автоматического управления», под ред. Б. Г. Доступова. М., «Машиностроение», 1970.
10. *Р. Л. Стратонович*. Принципы адаптивного приема. М., «Советское радио», 1973.

В. Н. Сорокин

Элементы кодовой структуры речи

Системы автоматического распознавания и понимания речи должны обеспечить человеку управление сложными автоматами, обеспечить доступ к автоматическим информационно-справочным системам, а в дальнейшем — помочь общению с искусственным интеллектом. Речевое общение человека с машиной эффективно только в том случае, когда система автоматического понимания речи достаточно надежна и не затрудняет процесс общения ошибками и переспросами. Существующие системы не обеспечивают такого уровня эффективности прежде всего потому, что речь сформировалась для общения между людьми, и для понимания ее важны такие категории, как мимика, жесты и, главным образом, смысл сообщения, которые пока недоступны автоматическому анализу. Другая причина состоит в том, что структура речевого сигнала чрезвычайно сложна, и в практических разработках до сих пор удавалось использовать лишь отдельные частные свойства речи. Только в последнее время начинают создаваться системы, учитывающие более широкий круг этих свойств: фонологические правила произношения, грамматические ограничения на структуру фраз и т. д.

Все более увеличивается потребность в синтезе сведений для создания модели структуры речевого сигнала, которая учитывала бы основные свойства речи и могла бы использоваться при раз-

работке алгоритмов принятия решений. Такая модель должна опираться на сведения о процессах речеобразования и восприятия, а методы ее исследования должны быть аналогичны методам, использующимся в теории кодирования. Это вытекает из основного назначения речи — передачи информации, причем в данном случае нас интересует лишь одна сторона этого сложного процесса — передача фонетической информации. Способность речевого канала поддерживать связь в условиях высокого уровня помех и искажений и возможность членения речевого потока на дискретные элементы (слова, слоги, фонемы) позволяют отнести речь к классу корректирующих кодов и применить для ее исследования методы, развитые в этой области. Вместе с тем имеются и некоторые отличия от обычной ситуации, рассматриваемой в теории кодирования. Эти отличия состоят в том, что пока не обнаружено никакой алгебраической структуры речевого кода и можно строить лишь процесс декодирования речи, тогда как изменять ее кодовые свойства нельзя.

Схема кодирования речевой информации устанавливается исследованием процессов речеобразования. Эта схема такова: слово является наименьшей смысловой единицей речи. Словесный образ мысли перекодируется в последовательность нейромоторных команд, управляющих движениями артикуляторных органов: языка, губ, небной занавески, голосовых связок и нижней челюсти. Кинематические и динамические свойства артикуляционного аппарата приводят к взаимному влиянию артикуляционных процессов, в результате чего наименьшей единицей артикуляции оказывается слог, и команды управления формируются для слога в целом [1, 2].

Для достижения необходимой точности артикуляции система управления включает в себя несколько цепей обратной связи: тактильную, проприоцептивную и акустическую.

Голосовой тракт представляет собой акустическую систему с распределенными параметрами. Поэтому каждое элементарное артикуляционное движение приводит к изменению многих акустических параметров речевого сигнала. Некоторый набор артикуляционных движений и состояний, выполняемых одновременно или последовательно, характеризует фонему — наименьшую слогоразличительную единицу речевого потока. Описание речевого сигнала в терминах фонем дает наиболее экономную запись сообщения.

Процессы кодирования речевой информации характеризуются большим разнообразием закономерностей, осложняющихся индивидуальными различиями произношения. Основные результаты в речевых исследованиях были получены путем выделения отдельных свойств и изучения их по возможности независимо от других. Вместе с продолжением этой важной работы необходимо и объединение уже известных закономерностей на основе более общего подхода к проблеме автоматического понимания речи.

Принцип «простоты». Результаты некоторых исследований свойств языка и восприятия показывают, что сложность процессов переработки информации играет важную роль в формировании структуры речи, например, закон Мандельброта [3], полученный из условия максимизации информации при ограниченной стоимости сообщения, удовлетворительно описывает распределение вероятностей слов и фонем (рис. 1). Эксперименты по восприятию речевых и неречевых стимулов [1, 4, 5] показывают, что пропускная способность мозга близка к 2 бит/символ, а оперативная память к 9 символам. При увеличении размерности пространства признаков разрешающая способность по каждому признаку падает до 2—3 градаций. Способность человека анализировать большие массивы информации может объясняться перекодированием [6], в процессе которого группе входных символов присваивается символ более высокого уровня, эти символы вновь группируются и т. д. до снижения объема информации до приемлемого уровня.

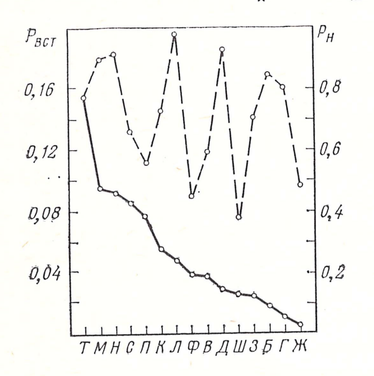


Рис. 1. Распределение вероятностей фонем

——— частота встречаемости звуков по [23];

— — — — — вероятность правильного приема звуков по [10]

Аналогичная идея многократной группировки символов привела к созданию каскадных кодов, отличающихся высокой исправляющей способностью и простотой декодирования [7].

Указания на то, что в формировании структуры речи сложность процессов генерирования и анализа речевых сигналов играет столь же большую роль, как и надежность речевой связи, приводят к некоторым важным следствиям. Прежде всего это дает основание для анализа структуры речевого кода методами, развитыми в теории кодирования, и позволяет применять в подходящих случаях приемы декодирования из этой теории. Отсюда также следует, что в целях экономии на каждом уровне речевого потока все имеющиеся связи могут не использоваться, и что некоторые уровни могут обладать даже нулевой исправляющей способностью, лишь обнаруживая ошибки.

Одно из практических следствий состоит в аналогии между речевым кодом на уровне слов и кодами Шеннона — Фано и Хафмена [8]. Распределение вероятностей, полученное максимизацией информации при ограниченной стоимости и предписывающее слову тем меньшую вероятность появления, чем оно «дороже», можно рассматривать как обращение принципа, используемого в этих кодах, по которому наиболее вероятным словам присваиваются более короткие коды, а наименее вероятным — длинные.

Как известно, коды Шеннона — Фано и Хафмена являются

неприводимыми (или префиксными) кодами, обладающими тем свойством, что никакое кодовое слово не является началом другого кодового слова, и это свойство позволяет выделять кодовые слова из непрерывной последовательности символов без каких-либо разделяющих сигналов. И хотя неприводимость не является обязательным признаком кодов, построенных по принципу «наиболее вероятные кодовые слова — наиболее короткие», можно предположить, что и реальный язык в известной степени подобен неприводимым кодам.

Проверка этого предположения проводилась на материале словаря русских литературных слов, содержащего 2500 наиболее часто встречающихся слов [9] и записанного в фонетической транскрипции. Слова попарно сравнивались друг с другом, причем совмещение слов происходило по их началам. Подсчет количества слов, которые служат началом других, показал, что всего таких слов нашлось 170 из 2500 (6,3%), и, таким образом, можно считать, что на уровне слов речь обладает свойством неприводимости. Общая вероятность появления слов-префиксов равна 0,278 за счет наиболее часто встречающихся одно- и двухбуквенных слов, без учета которых эта вероятность равна 0,097. Этот результат имеет очень важное практическое значение, поскольку все известные приемы сегментации слитного речевого потока на слова с использованием просодических характеристик (интонации, ударения), или опирающиеся на вероятности начальных и конечных звукосочетаний, не приводят к решению этой проблемы. Подобные косвенные оценки, безусловно, полезны, но они лишь дополняют механизм декодирования, основанный на префиксном свойстве слов. Важно также и то, что алгоритмы декодирования кодов Шеннона — Фано и Хафмена являются самокорректирующимися относительно границ между кодовыми словами, и это позволяет производить декодирование даже в том случае, когда начало слова по тем или иным причинам не было принято.

Акустические и артикуляционные признаки. Реальность существования артикуляционных признаков как различительных признаков фонем удостоверяется в результате анализа процесса речеобразования и некоторых особенностей процесса восприятия.

Некоторые артикуляционные признаки отчетливо проявляются в процессе речеобразования — это включение голосового источника, положение нёбной занавески, смычка и т. д. Эти признаки можно выявить и в процессе восприятия. Так, в матрицах переходов фонем, полученных при субъективном восприятии слогов ГСГ в условиях маскировки шумом, видно, что реализуются далеко не все возможные переходы, а ошибки совершаются за счет маскировки одного-двух артикуляционных признаков [10] (табл. 1).

Более того, известны эксперименты по кратковременному визуальному запоминанию слогов, в которых матрицы переходов букв в известной степени аналогичны матрицам переходов фонем,

Таблица 1

Матрица переходов фонем

	б	д	г	ж	з	в	л	м	н
б	76,1	13,3	6,5		0,4	0,9	0,7		
д	13	71,2	10,6		0,2	0,7	0,4		0,2
г	14,8	19,4	53,3		2,7	2,5	0,2		
ж	0,2	2,5	4	45,5	24,5	7,2	0,8		
з	2	11,5	8,9	4,2	41,9	10,4	2,3	0,1	0,3
в	3,9	5,1	12,2	0,7	4,5	48,2	5,7	0,4	0,3
л	0,1	0,3	0,3	0,1	0,5	0,6	96,4	0,5	0,6
м	0,1		0,1			0,1	2,4	65,7	29,2
н	0,1	0,1				0,1	2,9	10,8	82,7
п	0,1	0,2	0,2	0,1		0,3		0,1	
т			0,1		0,1				0,1
к								0,1	
ш			0,1					0,2	
с		0,1	0,1					0,1	0,1
ф			0,1						

так, как если бы происходило перекодирование зрительных образов в слуховые [11].

Хотя все артикуляционные признаки связаны с мышечными сокращениями в голосовом аппарате, проявление их в изменении геометрии голосового тракта и влияние на акустические характеристики речевого сигнала отличаются большим разнообразием и требуют специфических приемов анализа каждого признака.

Некоторые артикуляционные признаки по характеру сопровождающих их физических процессов тяготеют к двоичной градации (включено — выключено). Это признаки голосового источника, шумового источника, назальности, смычки, автоколебаний языка при артикуляции звука «*Р*». Каждому такому артикуляционному признаку соответствует ряд акустических признаков. Например, решение о наличии голосового возбуждения может быть принято и с помощью выделения импульсов основного тона, и анализом тонкой структуры спектра через кепстр, и анализом среднего наклона спектра. Признак назальности характеризуется в акустическом сигнале улучшением условий излучения на нижних частотах и появлением антирезонанса в области 700—1000 гц и т. д.

Другие артикуляционные признаки описывают различие форм голосового тракта, причем согласные звуки характеризуются локальным, концентрированным на небольшом участке изменением конфигурации тракта (так называемым «местом артикуляции»), тогда как гласные отличаются более общим характером изменений. Кроме того, имеются артикуляционные признаки, характеризующие движение артикуляторных органов. Все эти признаки имеют

при маскировке белым шумом

п	т	к	ш	с	ф	х	й	—	ч	ц
							0,8	1,3		
							2,9	0,7		
							5,9	1,2		
							13,7	1,6		
0,1							16,8	1,5		
	0,2						16,7	2,1		
							0,5	0,1		
0,1							1	1,3		
0,2							1,8	1,8		
56,3	14,2	6,3	0,7	5,7	11,2	2,6	0,4	1,4	0,2	
11,9	60,5	6,2	1	10,7	6,7	0,7	0,2	1,7	0,2	
11,4	19,7	46,1	1,3	9,4	7,9	2,2	0,1	1,4	0,5	
2,2	12,4	2,3	41,1	30,6	7	1,4	0,2	2	0,5	
4,6	19,7	2,8	3,4	56,7	8,2	1,5	0,4	1,4	0,6	0,4
16,4	12,8	7,9	1,6	15,2	37,6	6,5	0,5	1,1	0,3	

более сложную связь с акустическими характеристиками речевого сигнала и проявляются главным образом в положении и траекториях формант.

В [12] было показано, что в зависимости от того, приходится ли наибольшее сужение в акустической трубе на узел или пучность некоторой собственной функции этой трубы, происходит понижение или повышение соответствующей собственной частоты. Это свойство голосового тракта лежит в основе возможности принятия решений о качестве гласного на плоскости частот первых двух формант $F_1 - F_2$. Это же свойство используется и для определения места артикуляции по переходам частот формант, однако в этом случае принятие решений, как известно, затрудняется коартикуляцией — влиянием окружающих гласных звуков, на фоне которых происходит артикуляция согласного звука.

Анализ явления коартикуляции позволяет построить процедуры определения места артикуляции как по переходам частот формант [13—15], так и по динамике их амплитуд [16]. Наряду с акустическим анализом признака места артикуляции в последнее время большое внимание привлекает «анализ через синтез», позволяющий восстановить форму голосового тракта. Этот метод характеризуется большой помехоустойчивостью и дает возможность принимать решения о месте артикуляции по координате наибольшего сужения, но отличается большой трудоемкостью.

Таким образом, артикуляционные признаки по сложности анализа разбиваются на две группы: для одной группы достаточно использовать непосредственно акустические характеристики ре-

чевого сигнала, а для другой имеется необходимость в «анализе через синтез» с целью восстановления формы голосового тракта. Соответственно этим группам признаков существуют и две теории восприятия — акустическая и моторная. Согласно акустической теории, качество фонем определяется исключительно их акустическими различительными признаками. Моторная теория предполагает, что в процессе анализа речевого сигнала определяются параметры артикуляции, которые и служат первичным описанием фонетического состава речи. Обе теории подкрепляются экспериментальными данными, но ни одна из них не смогла занять доминирующего положения, что, по-видимому, отражает объективное различие в свойствах артикуляционных признаков.

Как видно из последующего, при хороших условиях речевой связи достаточно лишь акустического анализа, а при повышении уровня помех может появиться необходимость и в артикуляционном анализе. Физиологическое обоснование возможности артикуляционного анализа состоит в существовании акустической цепи обратной связи, которая решает задачу коррекции нейромоторных команд управления артикуляционными движениями по акустическим характеристикам текущего речевого сигнала [2].

Рассмотрим теперь возможную структуру процессов принятия решений относительно артикуляционных признаков, не касаясь конкретных алгоритмов их выделения непосредственно из акустического потока. Будем считать, что артикуляционные признаки, образующие код фонем, характеризуются постоянством значения, но интервалы времени, на которых они постоянны, для различных признаков различны и могут быть сдвинуты относительно друг друга. В этом случае препарирование речевого сигнала заключается в разделении его на последовательность неперекрывающихся сегментов, содержащих только по одному значению любого признака. Психофизические исследования показывают, что на этом уровне применимы методы статистической теории решений [17], поэтому после необходимой нормализации все вариации акустических признаков можно считать случайными, в том числе и ту изменчивость, которая порождается индивидуальными особенностями произношения и взаимным влиянием звуков. Такой подход оправдывается стремлением к максимальной простоте решающих правил и возможностью коррекции ошибок на фонемном и словесном уровнях.

Известно, что одному и тому же артикуляционному признаку соответствует ряд акустических явлений, и в результате этого возможны разнообразные алгоритмы выделения артикуляционных признаков. Если рассматривать акустический процесс речеобразования как многоканальную передачу информации о состоянии артикуляционного аппарата, то решение о наличии или отсутствии того или иного артикуляционного признака можно принимать методом накопления по множеству соответствующих ему акустических признаков

$$v_i = \frac{1}{n} \sum_{k=1}^{n} s_{ki}, \qquad (1)$$

где v_i — величина, пропорциональная некоторому артикуляционному признаку; s_{ki} — измеренные значения акустических признаков, нормированные таким образом, чтобы их можно было суммировать.

Такой способ обнаружения сигнала, как известно, может дать выигрыш в n раз в отношении сигнал/помеха, если значения помехи независимы. При выделении артикуляционных признаков этот выигрыш меньше n не только за счет коррелированности помех, но и за счет того, что на некоторые акустические признаки весьма сильно влияют и другие артикуляционные признаки, так что они характеризуют данный артикуляционный признак лишь с некоторой степенью принадлежности. Это явление учитывается с помощью весов α_{ki}, пропорциональных информативности акустического признака относительно i-го артикуляционного признака, и образования взвешенной суммы вместо (1):

$$v_i^* = \frac{1}{n} \sum_{k=1}^{n} \alpha_{ki} s_{ki}.$$

Количество артикуляционных признаков близко к десяти. В их число входят признаки звонкости, шумности, назальности, аналогичный признаку назальности признак разветвления воздушной струи при артикуляции звука «л», признак автоколебаний при артикуляции звука «р», признак степени раскрытия ротовой полости (иногда называемый признаком различения гласных от согласных) и признак места артикуляции. Согласно свойствам восприятия, каждый признак должен иметь не более двух-трех градаций, и это выполняется для всех признаков, кроме признака места артикуляции. По этому признаку можно выделить пять мест активной артикуляции, описывающих следующие группы фонем: губные, переднеязычные зубные, переднеязычные нёбные, среднеязычные и заднеязычные. Кинематика артикуляционного аппарата такова, что почти все эти движения могут выполняться независимо, так что фактически имеется пять двоичных признаков места артикуляции вместо одного с пятью градациями. Поэтому решающие правила для артикуляционных признаков могут быть двухальтернативными. Процесс принятия решений относительно некоторого сигнала обычно состоит в сравнении отношения апостериорных вероятностей (отношения правдоподобия) $\lambda = p(H_1 \mid v_i)/p(H_0 \mid v_i)$ с некоторым порогом λ_0, где H_1 и H_0—соответственно гипотезы о наличии или отсутствии сигнала.

Выбор порога λ_0 зависит от сведений о рассматриваемой задаче, в частности от стоимости ошибок и априорной вероятности появления признака. Если цены и априорные вероятности известны, то оптимальным является критерий минимального риска (крите-

рий Байеса); если известны только цены, применяют минимаксный критерий, гарантирующий, что риск не превзойдет некоторого уровня.

Стоимости ошибок на артикуляционно-фонемном уровне в настоящее время неизвестны, а априорные вероятности, установленные по средней частоте встречаемости, имеют значительную дисперсию, определяемую частотами слов в данном словаре и частотами появления фонем в некоторой позиции слова. Критерий Неймана — Пирсона не нуждается ни в ценах, ни в априорных вероятностях; он состоит в выборе такого λ_0, который бы минимизировал ошибки одного типа, например вероятность пропуска сигнала, равную

$$p_1 = \int_{\lambda_0}^{\infty} p(v)\, dv;$$

при заданной вероятности ошибки другого типа, например вероятности ложной тревоги

$$p_2 = \int_{0}^{\lambda_0} p(v)\, dv,$$

где $p(v)$ — вероятностное разделение данного признака.

Но в задаче распознавания артикуляционных признаков нет никаких оснований для предпочтения ошибок одного рода ошибкам другого рода, и поэтому критерий Неймана — Пирсона следовало бы применить дважды — для вычисления порога λ_{01} при условии $p_1 = \varepsilon$ и для вычисления порога λ_{02} при условии $p_2 = \varepsilon$. В общем случае $\lambda_{01} \neq \lambda_{02}$, и между ними остается зона неопределенности, при попадании в которую относительно измеряемой реализации не принимается никакого решения.

Артикуляционный признак сохраняет свое значение на некотором интервале времени τ, поэтому имеется возможность принимать решения не по однократному отсчету, а по целому ряду таких отсчетов. Простейший и наиболее часто применяемый способ вычисления функции правдоподобия состоит в усреднении $v_i(t)$ по времени; он адекватен гауссовской модели помех.

Распределение вероятностей $p(v)$, а с ним и пороги λ_{01} и λ_{02} зависят от уровня шумов. Поэтому в нестационарных условиях необходимо каким-либо образом производить оценку этого уровня. Относительно сигналов, попавших в зону неопределенности между порогами λ_{01} и λ_{02}, может быть либо принято решение о стирании, либо приписано значение правдоподобия [7]:

$$q(v) = \begin{cases} +1, & v \geqslant \Delta\lambda; \\ \dfrac{v}{\Delta\lambda}, & -\Delta\lambda \leqslant v \leqslant \Delta\lambda; \\ -1, & v \leqslant -\Delta\lambda; \end{cases} \qquad \Delta\lambda = \lambda_{02} - \lambda; \quad \lambda = \frac{\lambda_{02} - \lambda_{01}}{2}.$$

В дальнейших процессах принятия решений потребуются сведения о надежности того или иного артикуляционного признака. Их можно получить из психоакустических экспериментов с различными видами маскировки. В табл. 2 показана надежность выделения признаков звонкости, назальности, шумности и места артикуляции по результатам аудиторных испытаний при маскировке белым шумом для английского [18] и русского языков [10], при маскировке импульсной помехой с частотой следования около

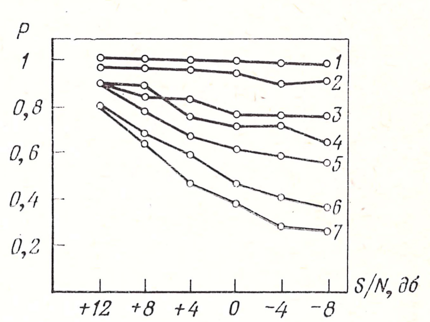


Рис. 2. Субъективная вероятность правильного приема артикуляционных признаков

1 — голосовой источник; *2* — назальность; *3* — переднеязычный; *4* — шумовой источник; *5* — губной; *6* — заднеязычный; *7* — переднеязычный — небный

120 гц, при полосовом ограничении 200 — 2500 гц и 200—1200 гц с отношением сигнал/шум +12 дб [18] и при клиппировании в полосе 7 кгц [15]. На рис. 2 показана зависимость надежности субъективного выделения артикуляционных признаков от отношения сигнал/шум. Из этого рисунка видно, что признаки звонкости и назальности наиболее надежны и мало зависят от уровня помех. В других условиях восприятия, где большую роль играет реверберация, наблюдается заметный процент переходов звонких и глухих звуков друг в друга [19].

Таблица 2

Субъективная вероятность правильного приема артикуляционных признаков: голосового источника (ГИ), назальности (Н), шумового источника (ШИ), места артикуляции (МА)

Маскировка и искажение	$p_{ГИ}$	$p_{Н}$	$p_{ШИ}$	$p_{МА}$
Белый шум, английский язык	0,97	0,946	0,815	0,707
Белый шум, русский язык	0,987	0,942	0,667	0,680
Импульсная помеха, русский язык	0,859	0,76	0,77	0,674
Полосовое ограничение 200—2500 гц, английский язык	0,981	1,00	0,893	0,814
Полосовое ограничение 200—1200 гц, английский язык	0,963	0,996	0,864	0,714
Клиппирование, русский язык	0,774÷0,964	0,785	0,931	0,699
Среднее	0,928	0,905	0,823	0,716

Фонемы. Определив фонему как совокупность артикуляционных признаков, можно подсчитать исправляющую способность речевого кода на уровне фонем. Если число несовпадающих артикуляционных признаков назвать кодовым расстоянием d, то условие обнаружения ошибок кратности Q_d есть $Q_d \leqslant d - 1$, а условие исправления ошибок кратности Q_c есть [20]: $2Q_c \leqslant d - 1$.

Составляя таблицу кодовых расстояний фонем и вычисляя по ней спектр взаимных расстояний (рис. 3), можно убедиться, что

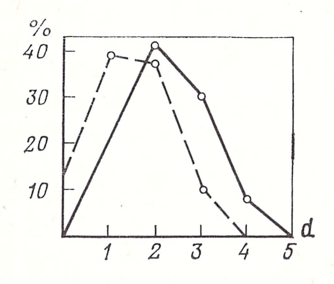


Рис. 3. Спектр кодовых расстояний фонем

——— полное описание;
— — — — — — без признака места артикуляции

на фонемном уровне потенциально возможно обнаружение около 75% и исправление около 37,2% одиночных ошибок, а число обнаруживаемых двойных ошибок меньше 10%. Стертые артикуляционные признаки, отношение правдоподобия которых попало в зону неопределенности, можно восстановить на уровне фонем, причем число восстанавливаемых признаков $Q_t \leqslant d - 1$ в данном случае равно 75%. Это указывает, что исправляющую способность фонемного кода выгоднее использовать для восстановления стертых признаков, а пороги в решающих правилах для артикуляционных признаков установить таким образом, чтобы увеличить число стираний и за этот счет повысить достоверность остальных решений. Как правило, стертый признак легче восстановить, чем исправить ошибку, потому что место стертого признака точно определено. Признаки места артикуляции обладают наименьшей помехоустойчивостью и наибольшей трудоемкостью.

На рис. 3 показано, как меняется спектр кодовых расстояний фонем, если признаки места артикуляции станут неразличимыми. При этом 12,4% фонем будут иметь одинаковые коды, обнаружение одиночных ошибок упадет до 47,7%, а исправление одиночных ошибок до 10,5%. Как видно, соотношение между числом восстанавливаемых стертых признаков и числом исправляемых ошибок здесь также в пользу первого. Стоимость генерирования каждого артикуляционного признака σ, вычисленная из закона Мандельброта, оказывается различной в зависимости от противопоставляемых звуков.

Вероятность правильного приема артикуляционных признаков в экспериментах по субъективному восприятию при маскировке также заметно различается. Это означает, что характеристики артикуляционных признаков при их генерировании и восприятии оказываются зависящими друг от друга. Очевидно, эта зависимость

частично является следствием уже обсуждавшегося свойства речевого сигнала — влияния различных артикуляционных признаков на один и тот же акустический признак. Декодирование фонем по зависимым признакам сложнее декодирования по независимым признакам, поэтому нужно оценить влияние зависимости признаков на надежность декодирования слов. С этой целью построим следующую математическую модель речевого кода на уровне фонем: при формировании слов последовательность фонем выбирается случайной, но зависящей от τ предыдущих фонем, и переходы фонем под действием помех независимы. В такой модели можно оценить верхнюю и нижнюю границы надежности декодирования слов при различных способах декодирования фонем с учетом и без учета зависимости артикуляционных признаков [1].

Используя теорему о кодировании, можно определить потенциальную вероятность ошибки $p(\varepsilon)$ при различении слов, обладающих кодовыми связями на τ фонем:

$$p(\varepsilon) \approx \exp(-E(R)\tau), \qquad (2)$$

где R — количество информации на фонему; $E(R)$ — функция надежности, вычисляемая по формуле $E(R) = \min\limits_{0 \leqslant x < \infty} [G(x) - x)R]$.

Здесь $G(x)$ — функция Галлагера [21], которая определяется соотношением

$$G(x) = -\ln \sum_{j=1}^{I} \left(\sum_{k=1}^{I} p_k p_{kj}^{1/(1+x)} \right)^{1+x},$$

где I — число фонем; p_k — вероятность появления k-й фонемы; $\| p_{kj} \|$ — матрица переходов фонем.

Графический способ вычисления $E(R)$ состоит в нахождении точки пересечения оси ординат, касательной к функции $(G(x) - xR)$, проведенной из точки, в которой $R = \mathrm{d}G/\mathrm{d}x$.

Матрицы переходов фонем $\| p_{kj} \|$ и матрицы переходов артикуляционных признаков были взяты из экспериментов по восприятию бессмысленных слогов типа гласный — согласный — гласный при разных отношениях сигнал/шум [10]. Матрицы переходов гласных звуков принимались единичными. Для оценки вероятности ошибки по (2) необходимо знать скорость передачи информации в речевом коде R_∞ и глубину кодовых связей. Скорость информации R_∞ принималась лежащей между 0,97 и 0,58 (в натах) [22], а глубина кодовых связей — равной длине слова n (при $n \leqslant 7$). Поэтому вместо (2) для расчетов использовалось

$$p(\varepsilon) = \sum_{\tau=1}^{n} p_\tau \exp(-E(R)\tau),$$

---

[1] Это исследование было проведено совместно с К. Ш. Зигангировым.

где p_τ — вероятность появления слова длиной τ, вычисленная по словарю [9] (суммарная вероятность появления слов длиной $n \leqslant 7$ в словаре близка к 0,74).

На рис. 4 показаны нижние границы функции надежности, вычисленные непосредственно по фонемам (как бы с полным учетом зависимости артикуляционных признаков) и по независимым артикуляционным признакам. По этим же графикам можно определить и верхнюю границу ошибок, называющуюся границей существования: $E(R) = \inf\limits_{0 \leqslant x \leqslant 1} [G(x) - xR]$. Эта граница гарантирует существование способа декодирования, обеспечивающего данную надежность.

В табл. 3 показаны верхние и нижние оценки вероятности ошибок на словесном уровне для всех длин слов ($1 \leqslant n \leqslant 7$) и без одно- и двухфонемных слов ($3 \leqslant n \leqslant 7$). Как видно, нижняя граница ошибок при декодировании по фонемам столь мала, что выходит за пределы точности исходных экспериментальных данных. На рис. 5 приведены верхние границы ошибок в сравнении с субъективными вероятностями ошибки слов, полученными для тех же отношений сигнал/шум при аудиторных испытаниях [23]. Из сравнения следует, что субъективные показатели ошибок находятся ближе к верхней границе ошибок, рассчитанных по независимым артикуляционным признакам, причем разница между разными способами декодирования фонем и субъективными вероятностями падает по мере снижения условия шума. Эти результаты можно трактовать таким образом, что при декодировании фонем может быть отдано предпочтение более простому анализу по независимым артикуляционным признакам, в особенности при умеренном уровне шума.

Учитывая малую исправляющую способность фонем и заметное различие в их помехоустойчивости (рис. 6), можно предположить, что основная коррекция ошибок в речевом коде происходит на вышележащих уровнях, и что чрезмерное усложнение анализа фонем неэффективно и нецелесообразно.

Слоги. Артикуляционные признаки, характеризующие фонему, во многих случаях не появляются одновременно, а распределены во времени. Например, временна́я структура согласного взрывного звука в окружении гласных такова: движение артикуляторного органа (формантный переход) — смычка (звонкая или глухая) — раскрытие смычки, имеющее одну или две акустические фазы: взрыв (может отсутствовать) и формантный переход. Поэтому анализ таких звуков должен производиться на интервале слога. Таким образом, слог оказывается не только единицей артикуляторной программы, но и единицей анализа. Коартикуляция и связанные с ней трудности сегментации речевого потока на фонемы привели в практике распознавания к использованию специфических единиц, охватывающих часть слога и называемых фрагментами или транземами. Недостатком этих единиц, так же как и сло-

Таблица 3

Границы надежности декодирования слов $n \leqslant 7$ и $3 \leqslant n \leqslant 7$

		Верхняя				Нижняя			
		S/N = – 8 *дб*	0	+ 8	+ 12	– 8	0	+ 8	+ 12
				$n \leqslant 7$					
Фонема	R 0,97	0,034	0,029	0,025	0,023	—	—	—	—
	0,58	0,021	9,019	0,016	0,014	—	—	—	—
Признак	0,97	0,113	0,106	0,066	0,051	0,085	0,072	0,034	0,011
	0,58	0,063	0,054	0,039	0,029	0,003	0,002	0,001	0,0001
				$3 \leqslant n \leqslant 7$					
Фонема	0,97	$3,4 \cdot 10^{-3}$	$2,3 \cdot 10^{-3}$	$1,5 \cdot 10^{-10}$	$1,2 \cdot 10^{-3}$	—	—	—	—
	0,58	$1,1 \cdot 10^{-3}$	$7,4 \cdot 10^{-4}$	$5,2 \cdot 10^{-4}$	$3,8 \cdot 10^{-4}$	—	—	—	—
Признак	0,97	$4,4 \cdot 10^{-2}$	$3,6 \cdot 10^{-2}$	$1,5 \cdot 10^{-2}$	$8,4 \cdot 10^{-3}$	$2,5 \cdot 10^{-2}$	$1,8 \cdot 10^{-2}$	$3,4 \cdot 10^{-3}$	$2 \cdot 10^{-4}$
	0,58	$1,3 \cdot 10^{-2}$	$9,9 \cdot 10^{-3}$	$4,7 \cdot 10^{-3}$	$2,5 \cdot 10^{-3}$	$5,3 \cdot 10^{-6}$	$4,3 \cdot 10^{-6}$	$3,4 \cdot 10^{-6}$	$2,9 \cdot 10^{-6}$

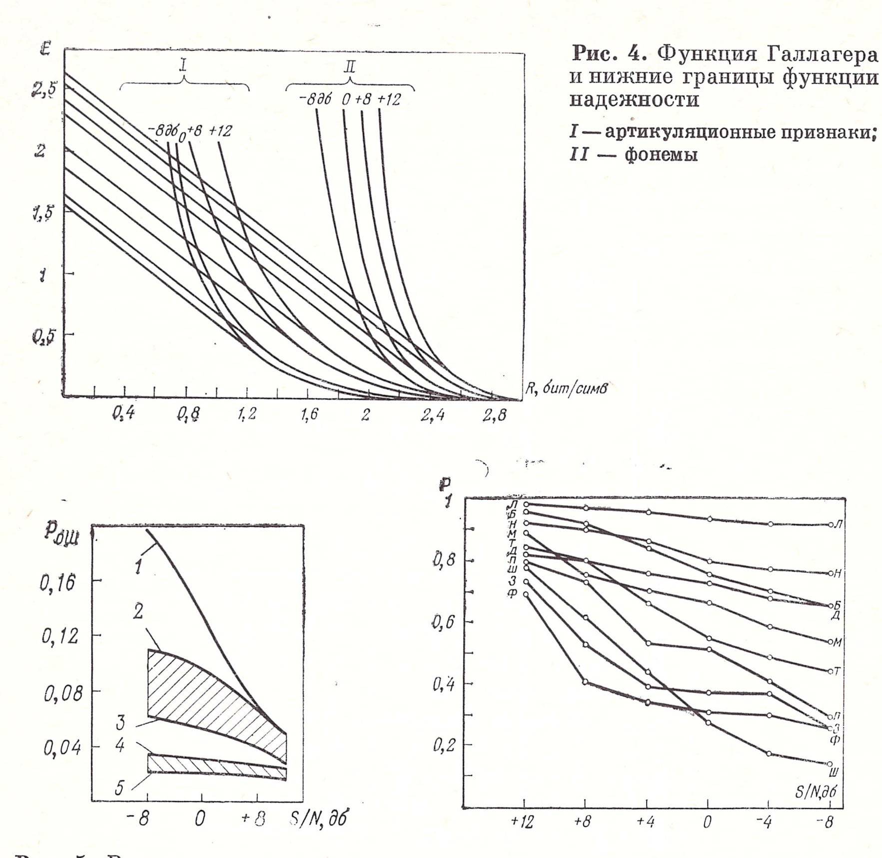


Рис. 4. Функция Галлагера и нижние границы функции надежности

I — артикуляционные признаки; *II* — фонемы

Рис. 5. Вероятность неправильного распознавания слов

1 — субъективная, по [23]; *2*, *3* — при декодировании фонем по независимым артикуляционным признакам; *4*, *5* — при декодировании слов непосредственно по фонемам. *2*, *4* — $R_{\infty} = 0{,}97$; *3*, *5* — $R_{\infty} = 0{,}58$

Рис. 6. Вероятность правильного приема некоторых фонем при маскировке белым шумом

гов, является то, что число их значительно (на один-два порядка) превышает число фонем, и описание речевого потока в их терминах весьма неэкономно.

Установлено, однако, что относительно малым числом слогов можно описать весьма большую часть речевых сообщений. Например, 100 фрагментов типа СГ охватывают 72% текста, содержащего более 1000 слов [24], и, следовательно, некоторые слоги могут быть включены в кодовую структуру речи.

Исправляющая способность слогов не исследована, но известны результаты аудиторных испытаний, связывающие фонемную слоговую и словесную разборчивость [23]. Из рис. 7 видно, что слоговая разборчивость всегда хуже фонемной, тогда как сло-

весная — всегда лучше слоговой. Даже если исправляющая способность слогов окажется ничтожно малой, их участок в кодовой структуре будет оправдан возможностью обнаружения физически неосуществимых последовательностей состояний. В каскадных кодах также иногда прибегают к введению каскада, обладающего лишь способностью к обнаружению ошибок.

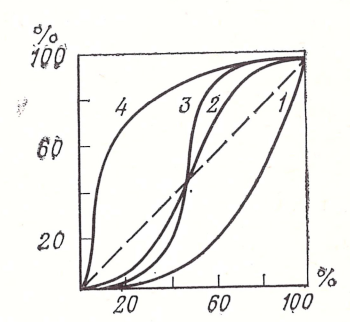


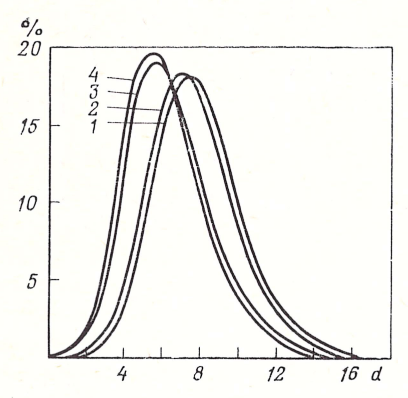
Рис. 7. Зависимость от фонемной разборчивости слоговой (*1*), словесной (*2*), фразовой (*3*) и словесной разборчивости от слоговой (*4*)

Слова. Фонетическая структура слов обладает значительно большей исправляющей способностью, чем фонемы. Эта исправляющая способность зависит от длины слов. Очевидно, что исправляющая способность однофонемных слов (предлогов и союзов) равна просто исправляющей способности фонем. Но средняя длины слов в русском языке близка к 6, и это позволяет обнаруживать и исправлять значительную долю ошибок даже в тех случаях, когда для описания фонем используются не все артикуляционные признаки, и некоторые фонемы становятся неразличимыми. Оценка спектров кодовых расстояний слов при различных способах вычеркивания артикуляционных признаков (объединения фонем в группы) производилась на словаре [9], записанном в фонетической транскрипции. В этом словаре кодовое расстояние между каждой парой слов определялось как минимальное расстояние по Хеммингу при всевозможных относительных сдвигах этих слов.

На рис. 8 условно в виде кривых показаны спектры кодовых расстояний для полных кодов фонем, а также для различных вариантов усеченных кодов. Один эксперимент был проведен при различении лишь трех мест артикуляции вместо пяти и четырех типов гласных (а, э; и, е, ы; о; у); исключены были также признаки автоколебательности и бокового прохода. В этом случае неразличимы, например, следующие звуки: (б, в), (д, р, л), (ж, з, и), (с, ш) и т. д. Другой эксперимент проводили при тех же условиях относительно согласных, но все гласные считались неразличимыми. Наконец, был полностью исключен признак места артикуляции (при сохранении различения согласных от гласных, также признаков звонкости, назальности и шумности).

На рис. 8 видно, что при использовании полных кодов фонем верхняя грань корректирующей способности слов составляет 99,9% обнаружения и 99,3% исправления одиночных ошибок,

96,3% обнаружения и 88,8% исправления двойных ошибок, 74,9% обнаружения и 57% исправления тройных ошибок и т. д., тогда как исключение признаков места артикуляции, автоколебательности и боковых проходов и полная неразличимость гласных понижают эти величины соответственно до 99% обнаружения и 96,4% исправления одиночных ошибок, 90,1% обнаружения и 80,4% исправления двойных ошибок, 66,7% обнаружения и 50,5% исправления тройных ошибок и т. д.



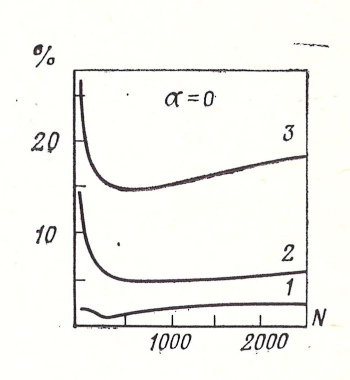


Рис. 8. Спектр кодовых расстояний слов

1 — полные коды фонем; *2* — три места артикуляции, четыре типа гласных, слияние признака автоколебательности и признака разветвлений переднеязычным местом артикуляции; *3* — то же, что и *2*; но все гласные не различимы; *4* — сохранены признаки звонкости, назальности, шумности, гласный — согласный

Рис. 9. Доля слов с одинаковыми кодами в зависимости от объема словаря

1 — соответствует *2*; *2* — соответствует *3*; *3* — соответствует *4* на рис. 8

Вместе с тем для усеченных кодов фонем некоторые слова начинают совпадать. Число слов, обладающих одинаковым кодом, в каждой группе невелико, а общее количество таких слов достигает 2,2% для первого способа усечения кодов, 5,6% — для второго способа и 17,9% — для третьего. На рис. 9 показана зависимость числа одинаковых слов от объема словаря.

Из этих расчетов следует, что при отсутствии помех подавляющее большинство слов может быть однозначно классифицировано с использованием лишь тех артикуляционных признаков (звонкость, шумность, назальность, признак «гласный — согласный»), которые выделяются непосредственно из акустического сигнала с помощью простых алгоритмов. По мере роста интенсивности помех роль признаков места артикуляции должна возрастать из-за необходимости обеспечения большего кодового расстояния между словами. Если процесс анализа речи организовать таким образом, что сначала с помощью акустического анализа выбирается группа слов, имеющих одинаковый код со словом на входе распознающе-

го устройства, а затем производится окончательная идентификация этого слова по полным кодам фонем, то среднее количество операций на декодированное слово уменьшается, и этот выигрыш тем больше, чем лучше условия связи.

При определенном уровне шумов потребуется использовать всю доступную избыточность, для чего необходимо включить механизм анализа через синтез. В этом случае акустические признаки послужат для направления поиска. Очевидно, что для выбора способа анализа нужно уметь измерять текущий уровень шумов.

Заключение. При построении модели речевого сигнала необходимо принимать во внимание сведения о роли фактора сложности в процессах речеобразования и восприятия, связанную с этим иерархическую структуру речи и возможность обнаружения и исправления ошибок. В результате этого для декодирования фонетической структуры (в противоположность оценке эмоционального и физического состояния человека по его речи) необязателен предельно точный акустический анализ речевого сигнала, который увеличивает разброс оцениваемых параметров и усложняет процесс декодирования. Вместе с тем должна быть сохранена необходимая точность для выделения акустических признаков, адекватных артикуляционным процессам речеобразования.

В число кодовых уровней речевого сигнала входят акустические и артикуляционные признаки, фонемы, слоги, слова и фразы. Каждый из этих уровней обладает в том или ином объеме способностью обнаруживать и исправлять ошибки, и это позволяет достичь необходимую надежность декодирования при использовании сравнительно простых алгоритмов. Сложность этих алгоритмов, число и вид используемых уровней зависят от условий в канале связи — в одних случаях можно ограничиться небольшим числом простых акустических признаков, в других случаях потребуется анализ через синтез, а при особенно тяжелых условиях необходимо использовать всю избыточность, заключенную в речевом сигнале, вплоть до семантического анализа.

К анализу речевого кода в ряде случаев применимы методы теории кодирования, что позволяет обнаружить свойства, важные с практической точки зрения, такие, как, например, свойство неприводимости, позволяющее выделять большинство слов из непрерывного потока речи. Из числа известных методов декодирования к речевому сигналу подходит только последовательное декодирование, причем специфика речи такова, что в процессе декодирования целесообразен параллельный анализ уровней речевого кода.

ЛИТЕРАТУРА

1. *Л. А. Чистович, В. А. Кожевников* и др. Речь. Артикуляция и восприятие. М.— Л., «Наука», 1965.
2. *В. Н. Сорокин.* К теории речеобразования.— Сб. «Речевое общение в автоматизированных системах». М., «Наука», 1975.
3. *Л. Бриллюэн.* Наука и теория информации. М., Физматгиз, 1960.

4. *I. Pollack, L. Ficks.* Information of multidimensional auditory displays.— JASA, 1954, v. 26.
5. *J. R. M. Hayes.* Memory span for several vocabularies as a function in a complex of vocabulary size.— Quart. Progr. Rep. Acoustic Lab. MIT. Cambridge, Mass., 1952.
6. *Дж. А. Миллер.* Магическое число семь плюс или минус два. О некоторых пределах нашей способности перерабатывать информацию.— Сб. Инженерная психология. М., «Прогресс», 1964.
7. *Д. Форни.* Каскадные коды. М., «Мир», 1970.
8. *Р. Фано.* Передача информации. Статистическая теория связи. М., «Мир», 1965.
9. *Э. Штейнфельдт.* Частотный словарь современного русского литературного языка. Таллин, 1961.
10. *В. Н. Сорокин, В. Н. Ложкин.* Анализ звуков речи методом маскировки.— Сб. «Теоретические и экспериментальные исследования в области структурной и прикладной лингвистики». Изд-во МГУ, 1973.
11. *W. A. Wickelgren.* Distinctive features and errors in shortter memory for English consonants.— JASA, 1966, v. 39.
12. *Дж. В. Стретт* (лорд Рэлей). Теория звука, т. 2. М., Физматгиз, 1955.
13. *S. E. G. Ohman.* Coarticulation in VCV utterances: spectrographic measurements.— JASA, 1966, v. 39, N 1.
14. *В. Н. Сорокин.* Об одной модели речеобразования.— Труды 6-го Всесоюзного семинара «Автоматическое распознавание слуховых образов». Таллин, 1972.
15. *Г. И. Цемель.* Опознавание речевых сигналов. М., «Наука», 1971.
16. *В. Н. Сорокин, В. С. Файн.* Признаки звонких взрывных.— АиТ, 1970, № 10.
17. *D. M. Green, J. A. Swets.* Signal detection theory and psychophysics. N. Y., 1966.
18. *G. A. Miller, P. E. Nicely.* An analysis of perceptual confusions among some English consonants.— JASA, 1955, v. 27, N 2.
19. *В. И. Бельтюков.* Об усвоении детьми звуков речи. М., 1964.
20. *У. Питерсон.* Коды, исправляющие ошибки. М., «Мир», 1964.
21. *К. Ш. Зигангиров.* Процедуры последовательного декодирования. М., «Связь», 1974.
22. *Р. Г. Пиотровский.* Информационные измерения печатного текста.— Сб. «Энтропия языка и статистика речи». Минск, 1966.
23. *Н. Б. Покровский.* Расчет и измерение разборчивости речи. М., Связьиздат, 1962.
24. *А. В. Книппер, В. А. Махонин.* К описанию речевых сигналов.— Сб. «Речевое общение в автоматизированных системах». М., «Наука», 1975.

Р. К. Потапова

О типологических особенностях слога

Одним из основных вопросов научного анализа является вопрос выбора объекта исследования и дальнейшей принципиальной стратификации его признаков. В связи с этим не менее важным представляется вопрос о том, какие объекты научного анализа следует считать *элементарными структурами* в рамках проводимого исследования.

Согласно теории познания, элементарная структура может быть рассмотрена как некоторая *целостность*, к уровню которой

должно быть сведено все многообразие данных исследования. Элементарная структура как объект исследования выступает в данном случае как своего рода атом, который оказывается в фокусе всего исследовательского построения. В этом контексте определенную значимость приобретают вопросы о *строении* и *функционировании* самой элементарной структуры. Всякая элементарная структура, рассматриваемая как целостность, должна обладать всеми признаками, присущими целостности как философскому понятию [2]. Понятие же целостности в этом аспекте весьма сложно по своему содержанию и далеко не исчерпывается чисто суммативным пониманием, при котором целостность сводится к сумме ее частей, что было характерным для понимания целостности с позиций механистического материализма [3].

Следует отметить, что применительно к исследованию речевого потока выбор основной элементарной структуры постоянно являлся важнейшим вопросом, находящимся в прямой зависимости от конечной цели исследования. Для лингвистических работ прикладного характера наметились в основном два пути, обусловленные природой объекта, выбранного в качестве опорной элементарной структуры. Сторонники одного подхода стоят на позициях вычленения в качестве опорной элементарной структуры звука (фонемы) и его субзвуковых (субфонемных) составляющих, сторонники другого подхода ориентируются на слог. Таким образом, на современном этапе развития проблемы автоматического распознавания речи можно зафиксировать наличие двух способов выбора опорной элементарной структуры как некоторой целостности в целях ее первичного распознавания.

Решение вопроса было бы не столь сложным, если бы каждую из названных элементарных структур (звук, слог) можно было бы рассматривать чисто суммативно. Однако, учитывая специфику слитной речи, следует на первое место при описании признаков целостного объекта поставить свойство *интегративности*. Подобный подход ведет к тому, что целостность будет характеризоваться новыми качествами и свойствами, не присущими ее отдельным составляющим, возникающими в результате взаимодействия этих составляющих в определенной системе связей. При этом одним из ведущих принципов следует считать принцип *субаддитивности*, согласно которому целое может быть меньше суммы его частей. Применительно к такой элементарной целостности, как слог, это означает, что, описывая признаки слога на артикуляторном, акустическом и перцептивном уровнях, можно предполагать наличие такого образования, которое несводимо к сумме его составляющих. При этом на слог как на интегративную целостность накладываются новые свойства, обусловленные не только природой его звуковых составляющих, но и условиями конкретной реализации, что в свою очередь может привести к образованию новых типов слога. Следует подчеркнуть, что в процессе научного познания и воспроизведения сложного объекта, каким является речевое высказыва-

ние, членение на составляющие может быть весьма разноплановым: параметрическим, сегментным, фонематическим, силлабическим, морфологическим, лексическим, синтаксическим, семантическим. На сложный объект не может быть наложено только одно теоретическое представление целостности. Анализируя сложный объект, исследователь каждый раз имеет дело с различными «срезами» объекта.

В литературе представлена точка зрения, согласно которой достаточно, например, располагать информацией о фонемной принадлежности составляющих слога, и вопрос об алфавите и типах слоговых структур вполне решен [9]. В принципе такой подход возможен, но тогда основной упор делается на фонотактику языка без дифференциации правил порождения объекта на фонетические и фонологические. При подобном подходе не только смешиваются понятия двух разных аспектов — речевого и языкового, но также и не учитывается интегративная сущность опорного вычленяемого объекта. Имеется и иной подход, при котором основополагающим при сегментации на слоги является параметрическая информация, как, например, информация об изменении уровня интенсивности сигнала [1]. Однако в данном случае основная ориентация на увеличение уровня интенсивности на участке, соответствующем гласному, не всегда правомерна. Известно, что реализация уровня интенсивности в слоге зависит от целого ряда переменных, таких, например, как функционирование долготы и краткости слогоносителя, изменение типа примыкания конечного согласного, временна́я организация слога, особый статус сонантов в ряде германских языков и т. д. В связи с последним фактором следует указать на то, что, например, в английском языке сонант может брать на себя в определенных случаях роль слогоносителя. Кроме того наблюдается, что в немецком и английском языках в закрытом слоге СГС с кратким гласным при сильном типе примыкания последующего сонанта фиксируется сплошь и рядом смещение максимума интенсивности от ядра — слогоносителя к конечному сонанту.

Таким образом, если для языков, где открытый тип слога является доминирующим, правомерна опора на элементарную структуру СГ или фрагмента, меньшего СГ, но охватывающего наиболее информативный участок слога [4, 5], то для германских языков выбор элементарной структуры на уровне слога далеко не универсален в рамках структуры СГ.

В качестве аргументации приведем некоторые данные, полученные нами в ходе дальнейшей разработки проблемы временно́й организации структуры речевого высказывания. Основное допущение на данном этапе исследования было сформулировано следующим образом: всякое речевое высказывание характеризуется определенным образом заданной и реализуемой в процессе речепроизводства временно́й программой. При этом все сегменты высказывания, являясь частью единого целого, подчинены определенной временно́й программе.

Исследования последних лет показали, что временна́я организация речи является сложным и многоплановым феноменом. Было найдено, что для различных языков характерен определенный тип временно́й организации слога. Известны попытки доказать, что при произнесении слова последнее программируется с точки зрения временно́й организации как единое целое, причем значимая зависимость существует между всеми сегментами, которые образуют слово.

Основная задача исследования заключалось в *поиске элементов* высказывания, между которыми существовала бы определенная значимая временна́я зависимость. Исследование проводилось на материале двух языков: немецкого и английского. Экспериментальный набор включал фразы, составленные с учетом целого ряда требований контекстуального характера ($n = 6$).

Экспериментальные фразы были начитаны на магнитофон МЭЗ-62 в студийных условиях носителями языка: немцами ($N_1 = 15$), англичанами ($N_2 = 10$). Значения длительности (в мс) для каждого анализируемого сегмента высказывания (звука, звукосочетания, слова, фразы) просчитывались по интонограммам, полученным с помощью интонографа И-67. Для выявления корреляции по длительности между анализируемыми сегментами высказывания в каждом конкретном случае определялся коэффициент корреляции $\bar{\rho}$:

$$\bar{\rho} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\sigma_x \sigma_y},$$

где x_i — абсолютное значение длительности, мс, каждого сегмента I; y_i — выборки II.

Исследование включало три серии проверки наличия — отсутствия корреляции по длительности между следующими элементами высказывания:

I. Двумя рядом стоящими звуками сло́ва в составе фразы по схеме: С — Ѓ и Ѓ — С.

II. Звукосочетаниями слова в составе фразы по схеме: СЃ — С (С) Г и СЃС — (С) Г.

III. Гласными слова в составе фразы по схеме: (С)Ѓ — (СС)Г.

Измерение коэффициента корреляции и проверка гипотезы H_0 о независимости связи по длительности между звуками С — Ѓ и Ѓ — С, взятыми последовательно в пределах каждого слова в составе анализируемых фраз, показали, что для английского и немецкого языков результаты *идентичны в плане общности тенденции*, но не идентичны в плане частоты встречаемости.

Для английского языка характерно наличие отрицательной корреляции по длительности между гласным и последующим со-

гласным в структуре слова независимо от позиций этого слова во фразе. Наличие отрицательной корреляции в данном случае означает, что изменение длительности гласного в сторону увеличения ведет к уменьшению длительности последующего согласного и наоборот. В сочетании С — Ѓ ни положительной, ни отрицательной корреляции по длительности регулярно почти не наблюдалось.

Для немецкого языка была выявлена отрицательная корреляция также для сочетания Ѓ — С, однако только в составе слов, занимающих конечную позицию во фразах. Корреляции по длительности между начальным согласным и последующим гласным обнаружить не удалось.

Проверка H_0 гипотезы о независимости связи в соответствии с задачей II серии исследования показала, что как для английского, так и для немецкого языков характерно наличие положительной корреляции по длительности между сегментами СЃС и (С)Г внутри слова. Корреляции по длительности между сегментами СЃ и С(С)Г в составе тех же слов не наблюдалось.

Сопоставление длительности гласных внутри структуры слова показало, что между гласными слова существует положительная корреляция. Причем применительно к немецкому языку следует констатировать наличие более регулярной картины в плане корреляции по длительности между гласными слова.

Резюмируя все выше изложенное, можно утверждать:

1) для обоих языков характерно наличие отрицательной корреляции по длительности между гласным и последующим согласным — ГС;

2) для обоих языков характерно наличие положительной корреляции по длительности между частями слов, взятыми по принципу: СЃС — (С)Г;

3) для обоих языков характерно наличие положительной корреляции по длительности между гласными в структуре слова.

На основании полученных данных логично далее предположить, что длительность элементов речевого высказывания характеризуется не только единой программой, но реализуется между элементами высказывания с разной степенью значимости и имеет иерархический характер.

Принимая во внимание иерархический (многоуровневый) характер зависимости по длительности между элементами высказывания, представляется вполне целесообразным в процессе исследования временно́й организации элементов высказывания пользоваться следующей схемой: вычленение в качестве объекта исследования временны́х связей: а) на уровне фраз; б) на уровне слога; в) на звуковом уровне.

На материале наших данных вычленение в качестве объекта исследования временны́х связей на сегментном уровне показало, что для английского и немецкого языков подтвердилась пра-

вомерность утверждения относительно наличия отрицательной корреляции по длительности между гласным и последующим согласным. Однако в отличие от данных, полученных ранее, длительность анализируемых сегментов рассматривалась не в изолированном звукосочетании типа ГС и не в структуре изолированного слова, а в *структуре фразы*. Дальнейшее варьирование сегментов в плане изменения характера комбинаторики последних показало, что для исследуемых языков в целом ряде случаев имеет место значимая временна́я зависимость между сегментами, скомбинированными по типу СЃС — (С)Г.

Что же касается корреляция по длительности между гласными слов во фразе, то здесь, очевидно, мы имеем дело скорее не с временно́й организацией сегментного уровня, непосредственно связанного с эффектом коартикуляции, а с временно́й организацией, связанной в большей степени с реализацией гласных по определенной временно́й программе в структуре слова, являющегося частью более сложного просодического целого — ритмического рисунка фразы.

Дальнейшее принятие во внимание масштаба фразы также привносит нечто своеобразное в картину временно́й зависимости частей внутри целого. Как показали результаты наших ранних исследований на материале других языков, в данном случае на первый план выдвигается влияние на длительность элементов позиционного фактора. Например, позиция конца фразы, характеризующаяся общей для различных языков тенденцией к увеличению длительности, представляет большую возможность для выявления временно́й зависимости между элементами высказывания.

На основании изложенных данных можно полагать, что временна́я структура речевого высказывания представляет собой не просто комбинацию значений длительности элементов, составляющих высказывание, а более сложную некоторую целостность, членящуюся на относительно автономные, внутренне связанные единства, подчиняющиеся как внутренним законам этих отдельных единств, так и законам всей целостности.

Полученные данные о наличии временно́й связи в закрытом слоге СГС в английском и немецком языках были подвергнуты проверке на перцептивном уровне [8]. В результате перцептивной сегментации и дальнейшей идентификации сигнала было обнаружено, что для носителей английского и немецкого языков одним из основных реальных типов распознаваемой слоговой структуры является закрытый слог СГС. Реализация конечного согласного при кратком слогоносителе содержит качественно-количественную информацию, позволяющую носителям языка правильно идентифицировать тип слога.

При описании алфавита слоговых структур, с нашей точки зрения, более корректно учитывать типологические особенности слога того или иного языка в целях наиболее адекватного описания [6, 7]. Определяя слог как элементарную сегментно-супрасегмент-

ную целостность, характеризующуюся набором объективных и субъективных признаков, можно полагать, что слог как результат действия фонотактических правил языка и как результат процесса речепроизводства проецируется на акустическую плоскость и благодаря этому может быть выделен чисто условно в терминах его акустических коррелятов:

1. Слог — единица языка. Сообщение в лингвистических единицах, представленное нервными образцами и нервными командами к мышцам. Здесь слог — часть языкового звука, границы которого определены правилами фонотактики языка.

2. Слог — единица речи: а) артикуляторные жесты. Здесь слог — последовательность артикуляторных жестов; б) речь, представленная акустическими сигналами. Здесь слог — сегментно-супрасегментная единица.

3. Слог — единица восприятия: а) слуховой анализ речевого сигнала. Здесь слог — последовательность сегментов и комплекс просодических признаков; б) сообщение в лингвистических единицах, расшифрованное из слуховых образцов. Здесь слог — часть языкового знака, совпадающая или несовпадающая с языковыми единицами других уровней.

Думается, что центральная проблема, на решение которой должен быть ориентирован один из этапов процесса распознавания слитной речи, заключается в поиске алгоритма распознавания сложного объекта на языке свойств его составляющих. Это можно представить в виде установленного однозначного соответствия между одним из подмножеств множества свойств частей целого и некоторым подмножеством множества свойств целого. Таким путем можно, очевидно, дать исчерпывающее описание всего множества свойств сложного целого. При этом необходимо опираться на лингвистические правила порождения высказывания в рамках той или иной языковой системы в целом. Весьма спорным представляется поиск универсальной программы выбора опорной элементарной структуры. Выбор последней детерминирован целым рядом факторов. Важно при этом анализировать и описывать признаки опорной элементарной структуры с позиций *интегративности*, учитывая влияние, обусловленное своеобразием реализации опорной элементарной структуры в слитной речи.

ЛИТЕРАТУРА

1. *Д. Аллен.* Система, понимающая речь, основанная на программном анализе предложения.— Труды 4-й Международной объединенной комиссии по искусственному интеллекту, ч. 5. М., 1975.
2. *И. В. Блауберг, Б. Г. Юдин.* Понятие целостности и его роль в научном познании. М., 1972.
3. *П. Гольбах.* Система природы. М., 1940.
4. *А. В. Книппер, В. С. Мирошников.* Текущее выделение фрагментов в речи.— Сб. «Автоматическое распознавание слуховых образов». Таллин, 1972.

5. *А. В. Книппер, В. А. Махонин.* Специфика описания речевого сигнала при распознавании.— Сб. «Речевое общение в автоматизированных системах». М., 1975.
6. *Р. К. Потапова.* Лингвистические и нелингвистические критерии слогоделения.— Сб. «Иностранные языки в высшей школе», вып. 2. Рига, 1975.
7. *Р. К. Потапова.* Некоторые вопросы сегментации речевого потока на слоги.— Сб. «Звуковая и семантическая структура языка». Фрунзе, 1975.
8. *Р. К. Потапова, Н. Г. Камышная.* Слог и его перцептивно-временные корреляты.— Вопросы языкознания, 1975, № 4.
9. *O. Fujimura.* Syllable as a unit of speech recognition.— JEEE Sympos. Speech Recognition. Carnegie Mellon Univ., 1974.

Ю. Н. Прохоров

Рекуррентное оценивание параметров речевых сигналов

Введение. Задача оценивания параметров речевых сигналов, характеризующих состояние артикуляторного аппарата при произношении звуков речи, по-видимому, может быть успешно решена с помощью методов прикладной статистики и теории случайных процессов. Такое направление исследований рассматривается в значительном количестве публикаций, посвященных различным аспектам проблемы анализа речевых сигналов [1—4]. Особенно большое внимание уделяется методам линейного предсказания, когда в качестве исходной модели сигнала используется конечно-разностное уравнение относительно невысокого порядка. Применение этой модели оказывается весьма полезным, так как допускает выбор оптимальных или близких к ним алгоритмов оценивания [3, 4], и, кроме того, полученные с помощью таких алгоритмов параметры достаточно хорошо отражают динамику артикуляторного аппарата в процессе произношения звуков речи [4]. В своем первоначальном виде указанные алгоритмы являлись довольно сложными процедурами, что затрудняло не только техническую реализацию, но и моделирование на ЭЦВМ [5]. Упрощение этих процедур может быть достигнуто посредством применения рекуррентных вычислительных приемов. К этому выводу практически одновременно пришли отечественные и зарубежные специалисты [6—10], после чего рекуррентные алгоритмы стали применяться как для решения систем автокорреляционных уравнений [10], так и для непосредственного оценивания параметров речи [6—9, 11, 12]. Дальнейшие исследования показали, что, по крайней мере во втором случае, не все простые и известные в других приложениях алгоритмы могут быть одинаково успешно применены в задаче анализа речи. Специфика сигнала, его ярко выраженная нестаци-

онарность приводят в отдельных случаях к резкому снижению скорости сходимости и/или к увеличению остаточной погрешности анализа.

Эти трудности могут быть преодолены с помощью улучшения исходной модели, а также посредством разработки (или выбора) рекуррентных алгоритмов, согласованных с особенностями структуры реального сигнала. Отмеченные направления развиваются многими специалистами [7, 11—13 и др.].

Настоящая работа посвящена вопросам построения рекуррентных алгоритмов с учетом структуры сигнала, а также возможностям снижения погрешности оценивания.

Модели сигнала. Будем предполагать далее, что речевой сигнал образован прохождением некоторого возбуждения ξ_t через линейную систему с дробно-рациональной передаточной функцией.

В этом случае справедливо следующее конечно-разностное уравнение:

$$x_t = \boldsymbol{\vartheta}^T \mathbf{x}_{t-m}^{t-1} + \mathbf{b}^T \xi_{t-l+1}^{t}, \tag{1}$$

где x_t — центрированная математическим ожиданием последовательность случайных отсчетов речевого сигнала; ξ_t — последовательность отсчетов возбуждения речевого тракта, которую удобно считать некоррелированной (когда это предположение неправомерно, характер ξ_t будет уточняться); t — нормированное дискретное время; $\boldsymbol{\vartheta}^T \triangleq [\vartheta_1, \ldots, \vartheta_m]$, $\boldsymbol{b}^T \triangleq [b_0, b_1, \ldots, b_{l-1}]$ — параметры модели; $\mathbf{x}_{t-m}^{t-1} \triangleq [x_{t-m}, \quad x_{t-m+1}, \ldots, x_{t-1}]^T$, $\xi_{t-l+1}^{t} \triangleq [\xi_{t-l+1}, \xi_{t-l+2}, \ldots, \xi_t]^T$; $\boldsymbol{\vartheta}$, $\boldsymbol{b}$ — векторы параметров, постоянных на интервале локального постоянства $T = 10 \div 20$ мс.

В частном случае $\boldsymbol{b} \equiv \mathrm{b}_0$ уравнение (1) приводит к модели линейного предсказания

$$x_t = \boldsymbol{\vartheta}^T \mathbf{x}_{t-m}^{t-1} + b_0 \xi_t, \tag{2}$$

где обычно полагают $m = 8 \div 12$, $b_0 = 1$.

С точки зрения анализа модель (1) является более сложной, чем (2), так как неизвестные параметры $\{b_i\}_{i=1}^{m}$ перемножаются с ненаблюдаемыми отсчетами последовательности ξ_t. В некоторых случаях, в частности на вокализованных участках незашумленного сигнала, это обстоятельство, по-видимому, не приводит к значительным затруднениям [11, 12].

Модели (1), (2) характерны тем, что при их использовании оценивание параметров следует проводить совместно. Определенный интерес вызывает вопрос о возможности построения модифицированной модели на основе (1), (2), допускающей последовательное (поочередное) оценивание параметров. С таких позиций в [14] была разработана многоэтапная модель сигнала, которая в частном случае совпала с моделью, предложенной F. Itakura,

Уравнения многоэтапной модели имеют вид

$$
\begin{aligned}
&{}^{+}\varepsilon_i(t) = \boldsymbol{a}_i^{T}\,{}^{-}\varepsilon_{i,\Pi}(t-(i-1)l-1) + u_t,\\
&{}^{+}\varepsilon_{i+1}(t) = {}^{+}\varepsilon_i(t) - \boldsymbol{a}_i^{T}\,{}^{-}\varepsilon_{i,\Pi}(t-(i-1)\,l-1),\\
&{}^{-}\varepsilon_{i+1}(t) = {}^{-}\varepsilon_i(t) - \boldsymbol{a}_i^{T}\,{}^{+}\varepsilon_{i,\Pi}(t+(i-1)l+1),
\end{aligned}
\qquad (3)
$$

${}^{+}\varepsilon_1(t) = x_t$, ${}^{-}\varepsilon_1(t-j) = x_{t-j}$, ${}^{-}\varepsilon_{i,\Pi}(t-j) = [{}^{-}\varepsilon_i(t-j), \ldots, {}^{-}\varepsilon_i(t-j-m)]^T$, ${}^{+}\varepsilon_{i,\Pi}(t) = [{}^{+}\varepsilon_i(t), \ldots, {}^{+}\varepsilon_i(t+m-1)]^T$, $\boldsymbol{a}_i = [a_{i1}, a_{i2}, \ldots, a_{il}, \ldots, 0]^T - m \times 1$, $i = 1, \ldots, m/l$; m/l — целое число и в упомянутом частном случае $l = 1$; a_i — вектор локально-постоянных параметров, которые могут быть пересчитаны в $\{\vartheta_i\}_{i=1}^{m}$.

Достоинством модели (3), кроме возможности последовательного оценивания, является связь коэффициентов a_i с цилиндрической аппроксимацией речевого тракта [13], что указывает на ее хорошую согласованность со структурой реального сигнала.

Предположение о локальном постоянстве параметров может быть снято, если задать характер их изменения во времени. Так, например, естественным обобщением является переход к марковским параметрам в (3):

$$
a_i(t) = F_i a_i(t-1) + G_i \eta_i(t), \qquad (4)
$$

где коэффициенты F_i, G_i известны априори; $\eta_i(t)$ — некоррелированная последовательность случайных величин.

Такие общие ситуации рассмотрены в [14, 15].

Алгоритмы оценивания на основе оптимальной рекуррентной фильтрации. Оценивание параметров модели линейного предсказания совпадает с задачей идентификации линейной системы, входной и выходной сигналы которой связаны уравнением (2). Субоптимальный рекуррентный алгоритм оценивания (идентификации) при гауссовской последовательности ξ_t и некоторых ограничениях на начальные условия (x_0, ξ_0) может быть получен из общих уравнений фильтрации условно-гауссовских последовательностей [16]:

$$
\mathbf{m}_{t+1} = \mathbf{m}_t + \gamma_t\,[b_0^2 + (\mathbf{x}_{t+1-m}^{t})^T \gamma_t \mathbf{x}_{t+1-m}^{t}]^{-1}\,[x_{t+1} - \mathbf{m}_t^T \mathbf{x}_{t+1-m}^{t}]\,\mathbf{x}_{t+1-m}^{t},
$$

$$
\gamma_{t+1} = \gamma_t - \gamma_t \mathbf{x}_{t+1-m}^{t}\,[b_0^2 + (\mathbf{x}_{t+1-m}^{t})^T \gamma_t \mathbf{x}_{t+1-m}^{t}]^{-1}\,(\mathbf{x}_{t+1-m}^{t})^T \gamma_t^T, \qquad (5)
$$

где $\boldsymbol{m}_t$ — оценка вектора ϑ в момент времени t; b_0 — может быть принят равным единице.

В отличие от оптимального алгоритма в (5) предполагается относительная произвольность начальных условий $(\boldsymbol{m}_0, \gamma_0)$. Применение процедур, подобных (5), описано в ряде работ зарубежных авторов [3, 4, 12]. Однако сложность алгоритма, которая обусловлена необходимостью вычисления матрицы γ_t, в значительной мере ограничивает его практическое использование. Существенное упрощение (5) при сохранении основных достоинств дости-

гается изменением второго уравнения

$$\mathbf{m}_{t+1} = \mathbf{m}_t + \varkappa_{t+1} [x_{t+1} - \mathbf{m}_t^T \mathbf{x}_{t+1-m}^t] \mathbf{x}_{t+1-m}^t, \tag{6}$$

$$\varkappa_{t+1} = \left[1 + \sum_{i=0}^{t+1} \| \mathbf{x}_{i-m+1}^i \|^2\right]^{-1} \sim \left[1 + m \sum_{i=0}^{t+1} x_i^2\right]^{-1}, \tag{7}$$

где $\varkappa_{t+1}$ — константа сходимости; $\| \mathbf{x}_{i-m+1}^i \|^2 = \sum_{l=0}^{m-1} x_{i-l}^2$.

Нетрудно убедиться, что {(6), (7)} совпадает с (5) в случае одномерного ϑ.

Эвристический способ построения уравнения (7) не гарантирует сходимости оценок m_t к истинным параметрам, поэтому возникает необходимость теоретического анализа свойств предложенного алгоритма. Такой анализ со значительным сокращением приведен в приложении 1, где возбуждение ξ_t полагается некоррелированной шумовой последовательностью. В реальном случае, когда оцениваются параметры вокализованного участка речи, это предположение не выполняется. Однако с помощью простых рассуждений можно убедиться в правомерности вывода о сходимости алгоритма и для реального сигнала.

Действительно, на интервалах между импульсами основного тона модель (2) вполне приемлема, так как отличие речи от процесса авторегрессии можно учесть введением эквивалентного некоррелированного шума, действующего на входе системы (2) и образующего вместе с шумовым возбуждением эквивалентный входной сигнал ξ_t. Тогда приведенные в Приложении 1 результаты дают основание утверждать, что на указанных интервалах сходимость имеет место. Полагая импульс основного тона достаточно коротким (это всегда можно делать, относя отличие реального импульса от предполагаемого к влиянию системы (2) с параметрами $\{\vartheta_i\}_{i=1}^m$), рассмотрим алгоритм {(6), (7)} в момент, совпадающий с временно́й координатой импульса. Заметим, что величина $| x_t |$ с приходом импульса возбуждения резко возрастает и, таким образом, константа $\varkappa_t$ резко убывает, компенсируя локальное возрастание уклонения $[x_{t+1} - \boldsymbol{m}_t^T \boldsymbol{x}_{t+1-m}^t]$. Следовательно, влияние наблюдений x_t в отмеченные моменты времени на процедуру оценивания должно быть незначительным.

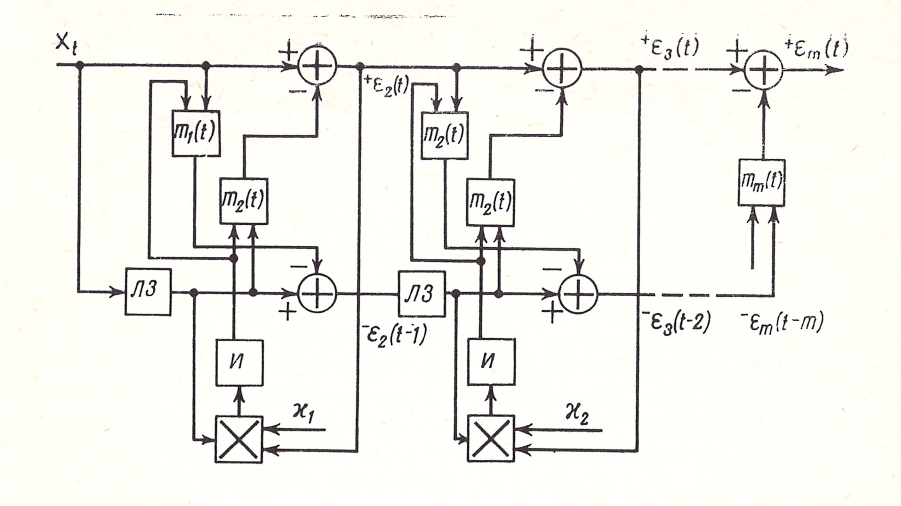
Несомненно, что приведенные рассуждения нуждаются в экспериментальном подтверждении.

Алгоритм {(6), (7)} применительно к многоэтапной модели с $l = 1$ принимает форму

$$m_i (t + 1) = m_i (t) + \varkappa_i (t + 1) {}^{+}\varepsilon_{i+1} (t + 1) {}^{-}\varepsilon_i (t - i + 1), \tag{8}$$

$$\varkappa_i (t) = \left[1 + \sum_{l=i}^{t} {}^{-}\varepsilon_i^2 (l - i)\right]^{-1}, \quad i = 1, \ldots, m, \tag{9}$$

где $m_i (t)$ — оценка a_i в момент t; ${}^{+}\varepsilon_{i+1} (t + 1) = {}^{+}\varepsilon_i (t + 1) - m_i (t) {}^{-}\varepsilon_i (t + 1 - i)$.



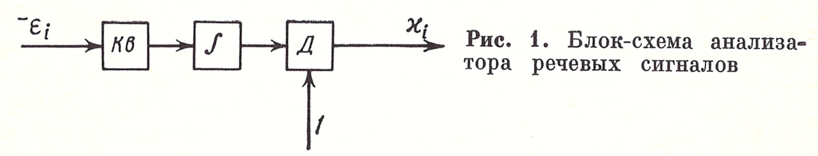


Рис. 1. Блок-схема анализатора речевых сигналов

Процедура оценивания {(8), (9)} может быть организована двумя способами: последовательно, когда сначала оценивается первый параметр a_1 по некоторому сегменту x_k, $k = t, t + 1, \ldots, t + N$, затем a_2 по сегменту x_k, $k = t + N + 1, \ldots, t + 2N$ и т. д., и параллельно, когда все параметры оцениваются одновременно с обязательной проверкой условия $|m_i(t)| \leqslant 1$ в каждый момент времени и установкой $m_j(t) = \operatorname{sgn} m_j(t)$, если это условие не выполняется для какого-то j. Последовательное оценивание реализуется в структурной схеме рис. 1, где блоки $m_i(t)$, в которых осуществляется усиление сигнала с весом $m_i(t)$, управляются выходными сигналами интеграторов И.

Алгоритм (6) при различном выборе $\varkappa_t$ уже применялся в задаче анализа речи [6, 7, 9], и в связи с этим в [14] отмечаются свойства таких алгоритмов в сравнении с {(6), (7)}.

Отметим попутно несколько эвристических приемов ускорения сходимости предложенного алгоритма. Так же, как и в обычной стохастической аппроксимации, можно ввести контроль знака погрешности $[x_t - \boldsymbol{m}_t^T \boldsymbol{x}_{t-m}^{t-1}]$. Тогда изменение константы $\varkappa_t$ следует производить лишь после изменения знака этого уклонения.

Допустимо также многократное прохождение по одному сегменту реализации в ускоренном масштабе времени.

Рекуррентная интерполяция в задаче анализа речи. Алгоритмы оценивания (5) — (9) хорошо приспособлены к последовательному поступлению наблюдений x_t на вход анализатора. При этом уст-

ройство функционирует в реальном масштабе времени. Вместе с тем на практике часто имеется возможность регистрировать фрагмент реализации сигнала в устройстве памяти. Так, в случае анализа на ЭЦВМ такой фрагмент равен массиву, считываемому с устройств внешней памяти в ОЗУ, и может составлять около $4 \cdot 10^3$ отсчетов (~0,4 мс при интервале дискретизации 0,1 мс). В такой ситуации появляется возможность улучшить точность оценивания за счет эффекта накопления.

Сформулируем задачу более строго. Пусть имеется фрагмент сигнала длительностью T, а в качестве исходной принята многоэтапная модель с $l = 1$. Поскольку оценка первого параметра не зависит от оценок последующих параметров, рассмотрим оценивание лишь на первом этапе. Исходные уравнения этого этапа выпишем на основе (3), (4) в виде (2.1), (2.2) — см. приложение 2. Формально (2.2) неправомерно, так как значения $a_1(t)$ не должны выходить за пределы $[-1, +1]$, однако при достаточно малой дисперсии $a_1(t)$ можно с целью упрощения результатов применить введенное уравнение.

Необходимо оценить $a_1(t)$, $t \in [0, T]$ оптимальным в рамках (2.1), (2.2) образом, используя весь фрагмент реализации x_t. Подобная задача решается методами теории оптимальной интерполяции [16], однако реализация известных процедур вызывает затруднения. В связи с этим в приложении 2 получено новое уравнение оптимальной интерполяции для (2.1), (2.2), при котором процедура оценивания сводится к двукратной обработке фрагмента $[0, T]$ в прямом и обращенном времени с помощью одного рекуррентного алгорита, вытекающего из уравнений оптимальной фильтрации. В квазиоптимальном случае процедура интерполяции описывается выражениями

$$d^{\pm}m_t/dt = F_{01} + F_1 \cdot {}^{\pm}m_t + 2\,{}^{\pm}D_t\,[x_t - {}^{\pm}m_t x_{t\mp\tau}]\,x_{t\mp\tau},$$
$$d^{\pm}D_t/dt = -2F_1 \cdot {}^{\pm}D_t + G^2/2 - 2\,{}^{\pm}D_t^2 x_{t\mp\tau}^2, \qquad (10)$$

где ${}^{\pm}m_t$, ${}^{\pm}D_t$ — оценки и их дисперсии в прямом времени (знак «+»; t возрастает от 0 до T) и обращенном времени (знак «—»; t убывает от T до 0).

Окончательная оценка m_t параметра a_{1t} и дисперсия D_t удовлетворяют соотношениям

$$m_t/D_t = {}^{+}m_t/{}^{+}D_t + {}^{-}m_t/{}^{-}D_t - x_t x_{t-\tau}/B_0^2 - F_{01}/D_a,$$
$$1/D_t = 1/{}^{+}D_t + 1/{}^{-}D_t - x_{t-\tau}^2/B_0^2 - 1/D_a, \qquad (11)$$

где D_a — дисперсия $a_1(t)$.

Таким образом, используя приведенные уравнения, можно оценить траекторию параметра $a_1(t)$, $t \in [0, T]$ и затем перейти к оцениванию параметров последующих этапов с помощью аналогичной процедуры, либо посредством алгоритма {(8), (9)}. При локальном постоянстве параметров модели, т. е. $F_{01} = F_1 = G_1 =$

$=0$, $a_1(t)=a_1$ и малом T, уравнения для оценок в дискретном времени получают вид

$$
\begin{aligned}
{}^{+}m_{t+1} &= {}^{+}m_t + \left[1+\sum_{l=1}^{t} x_l^2\right]^{-1}\left[x_t - {}^{+}m_t x_{t-1}\right] x_{t-1},\\
{}^{-}m_t &= {}^{-}m_{t+1} + \left[1+\sum_{l=T}^{t+1} x_l^2\right]^{-1}\left[x_t - {}^{-}m_t x_{t+1}\right] x_{t+1}.
\end{aligned}
\tag{12}
$$

При больших t ${}^{+}\varkappa_t \cong {}^{-}\varkappa_t$.

Отличительной чертой $\{(11),\ (12)\}$ по сравнению с $\{(8),\ (9)\}$ является выравнивание влияния наблюдений x_t, $t\in[0,\ T]$, на формирование оценки. Действительно, в случае $\{(8),\ (9)\}$ наблюдения в левой части интервала имеют больший вес, чем остальные, из-за монотонного убывания $\varkappa_t$.

Экспериментальное исследование алгоритмов рекуррентного оценивания. Экспериментальная проверка алгоритмов $\{(6),\ (7)\}$ и $\{(8),\ (9)\}$ была выполнена с помощью моделирования на ЭВМ «Минск-22». В качестве исходного материала были использованы шесть слов, произнесенных мужским и женским голосами (м.г.) и (ж.г.). Речевой сигнал был получен с помощью электродинамического микрофона и представлен в дискретной форме со следующими параметрами: частота дискретизации — 10 кгц, ширина спектра — 4 кгц, количество уровней квантования — 2^8.

Обработка сигнала при моделировании последовательного оценивания согласно $\{(8),\ (9)\}$ была организована следующим образом: реализация разбивалась на сегменты длительностью 12,8 мс, на которых проводилось оценивание; в качестве начальных условий на каждом сегменте принимались оценки, полученные на предыдущем сегменте; начальные условия на первом сегменте были нулевыми; количество параметров было равно 10.

На рис. 2, *а* приведены нормированные корреляционные функции процесса ${}^{+}\varepsilon_{10}(t)$ с величиной дисперсии σ_ε^2 для слова «миллион» ж.г. ($\sigma_\varepsilon^2 = 0{,}132$; $r_\varepsilon(0)=1$) и на рис. 2, *б* — «логарифм» м.г. ($\sigma_\varepsilon^2 = 0{,}14$; $r_\varepsilon(0)=1$). На рис. 3 показаны сегменты реализаций ${}^{+}\varepsilon_{10}(t)$ и x_t для слова «миллион» м. г. Следует уточнить, что столь четкое проявление импульсов основного тона в реализации ${}^{+}\varepsilon_{10}(t)$ наблюдается не на всех сегментах. Динамические траектории оценок $\{m_i(t)\}_{i=1}^{5}$ при наличии белого шума наблюдения с отношением шум/сигнал 0,1 для слова «логарифм» м.г. показаны на рис. 4, где t — количество итераций алгоритма оценивания.

Траектории оценок m_i и текущей нормированной дисперсии $\bar{\varepsilon}_0^2 = {}^{+}\tilde{\varepsilon}_{10}^2/\tilde{x}_t^2$ для слова «миллион» м.г. иллюстрируется рис. 5. С помощью двухстороннего порогового ограничения ${}^{+}\varepsilon_{10}(t)$ для полностью озвученного отрезка слова «миллион» м.г. были выделены координаты и амплитудные значения импульсов возбуждения, которые использовались при синтезе исходного сигнала по оценкам параметров. Нормированные корреляционные функции

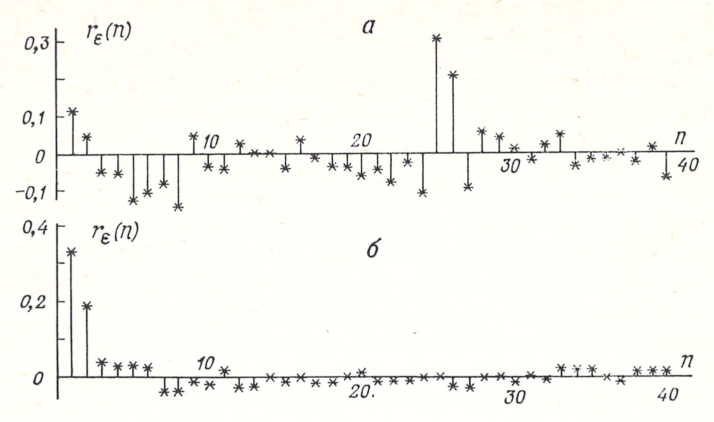
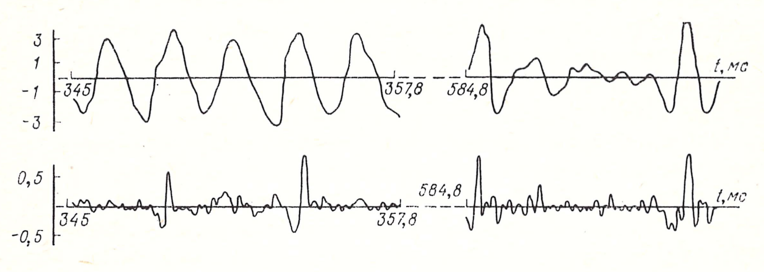


Рис. 2. Корреляционные функции процесса $^{+}\varepsilon_{10}(t)$



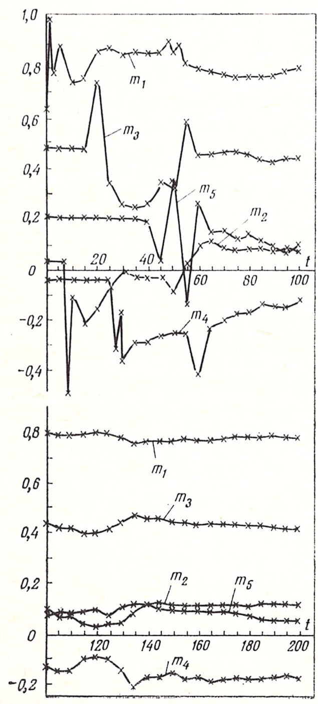


Рис. 3. Фрагменты реализаций сигнала и уклонения $^{+}\varepsilon_{10}(t)$

Рис. 4. Динамические траектории оценок $\{m_i(t)\}_{i=1}^{5}$

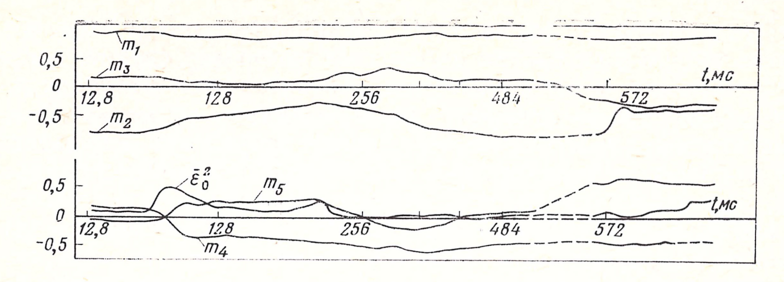


Рис. 5. Траектории оценок m_i для слова «миллион»

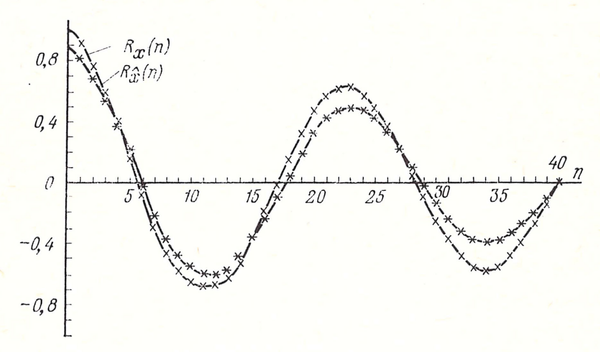


Рис. 6. Корреляционные функции исходного и синтезированного сигналов

исходного и синтезированного сигналов приведены на рис. 6. Исследование квантования оценок по остаточному процессу $^+\varepsilon_{10}(t)$ показало возможность квантования на $2^5 - 2^6$ уровней без существенного изменения характеристик $^+\varepsilon_{10}(t)$.

Результаты применения алгоритма $\{(6), (7)\}$ оказались несколько хуже и из-за ограниченности объема настоящей работы не приводятся.

Сделаем несколько замечаний:

1. Близость корреляционных функций $^+\varepsilon_{10}(t)$ к корреляционной функции некоррелированной последовательности указывает на достаточно высокую точность оценивания и хорошую согласованность модели с реальным сигналом.

2. Характер реализации $^+\varepsilon_{10}(t)$ подтверждает замечание 1, иллюстрируя в отличие от усредненных корреляционных функций текущее качество оценивания.

3. Точность оценивания параметров речи, произнесенной женским голосом, несколько хуже, чем мужским (см. рис. 2).

4. Наличие импульсов основного тона и шума наблюдения не приводит к расходимости алгоритма, что подтверждает приведенные ранее рассуждения. Вместе с тем в начале процедуры оценивания импульсы возбуждения могут влиять на формирование оценок (см. рис. 4, $t = 50$, $t = 135$).

Проведенная экспериментальная проверка разработанных на основе теории фильтрации алгоритмов позволяет сделать вывод о целесообразности их применения в задачах синтетической телефонии или автоматического распознавания слуховых образов. В последнем случае получаемые оценки могут быть использованы как для расчета формантных параметров, так и в качестве первичного описания сигнала, что, вероятно, более привлекательно.

Несомненный интерес представляет экспериментальное исследование рекуррентных алгоритмов интерполяции, которое, по-видимому, составит содержание последующей работы в этом направлении.

Приложение 1. Для краткости рассмотрим одномерный случай. Алгоритм оценивания ϑ при каждом фиксированном t можно представить в форме

$$m_\nu^{(t)} = m_{\nu-1}^{(t)} + \varkappa_\nu (x_\nu - m_{\nu-1}^{(t)} x_{\nu-1}) x_{\nu-1}, \tag{1.1}$$

$$\varkappa_\nu = \left[1 + \sum_{i=-[t^\alpha]+1}^{\nu} x_i^2\right]^{-1}, \tag{1.2}$$

где $x_n = \vartheta x_{n-1} + \xi_n$, $0 < \alpha < 1$, $0 \leqslant \nu \leqslant t \geqslant 0$, $m_0^t = 0$, $-\infty < n < \infty$, $[t^\alpha]$ — целая часть t^α.

Введем следующие ограничения:

1. Последовательность ξ_t некоррелирована и $E\xi_t = 0$, $E\xi_t^2 = \sigma_\xi^2 < \infty$, $E\xi_t^3 < \infty$. 2. Случайная последовательность x_t стационарна и удовлетворяет условию равномерно сильного перемешивания (р.с.п.). 3. $Ex_t^2 < \infty$, $|\vartheta| < 1$, $Ex_t = 0$, $Ex_t^3 < \infty$. Тогда справедлива следующая теорема.

Теорема.

1. При выполнении условий 1, 3 оценка $m_\nu^{(t)}$, определяемая из (1.1), (1.2), сходится в среднеквадратическом к ϑ:

$$\underset{t\to\infty}{\text{l.i.m.}}\, m_t^{(t)} = \vartheta. \tag{1.3}$$

2. Пусть имеют место ограничения 1—3. Тогда оценка $m_\nu^{(t)}$ асимптотически нормальна

$$L\,[\sqrt{t + [t^\alpha]}\,(\vartheta - m_t^{(t)})] \xrightarrow[t\to\infty]{} \frac{\sigma_x}{\sigma_\xi \sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{z^2}{2}\frac{\sigma_x^2}{\sigma_\xi^2}\right) dz, \tag{1.4}$$

где $L(z)$ — функция распределения величины z.

Докажем сформулированную теорему. Введем текущее отклонение $c_\nu^{(t)} = \vartheta - m_\nu^{(t)}$. Следовательно,

$$c_\nu^{(t)} = [1 - \varkappa_\nu x_{\nu-1}^2]\, c_{\nu-1}^{(t)} - \varkappa_\nu x_{\nu-1}\xi_\nu \Rightarrow c_\nu^{(t)} = \prod_{k=1}^{\nu} [1 - \varkappa_k x_{k-1}^2]\, c_0 -$$

$$- \sum_{k=1}^{\nu} \varkappa_k x_{k-1} \xi_k \prod_{l=k+1}^{\nu} [1 - \varkappa_l x_{l-1}^2].$$

Далее будет полезна следующая лемма, которую приведем без доказательства.

Лемма.

1. В случае ограничений 1—3 имеют место соотношения

$$\varkappa_t \xrightarrow[t\to\infty]{\text{с. к.}} 1/[1+(t+[t^\alpha])\,\sigma_x^2],$$

$$\chi_t = (1 - x_{t-1}^2/[1+(t+[t^\alpha])\,\sigma_x^2]) \xrightarrow[t\to\infty]{\text{п. н.}} 1 - \sigma_x^2/[1+(t+[t^\alpha])\,\sigma_x^2].$$

2. Последовательность $X_\mu = x_{\mu-1}\xi_\mu$, $-\infty < \mu < \infty$ удовлетворяет условию р.с.п. Используя лемму, вычислим следующий предел

$$\lim_{t\to\infty} E\left|c_t^{(t)} - \frac{1}{(t+[t^\alpha])\,\sigma_x^2}\sum_{k=1}^{t} x_{k-1}\xi_k\right|^2.$$

$$1.\quad \lim_{t\to\infty} E\,|\,c_t^{(t)}\,|^2 = \lim_{t\to\infty} E\left\{\prod_{k=0}^{t}\chi_k^2 c_0^2 + \right.$$

$$+\sum_{i=0}^{t}\sum_{j=0}^{t}\varkappa_i\varkappa_j x_{i-1}x_{j-1}\xi_i\xi_j \prod_{l=i+1}^{t}\chi_l \prod_{m=j+1}^{t}\chi_m -$$

$$\left. -2\prod_{k=0}^{t}\chi_k c_0 \sum_{m=0}^{t}\varkappa_m x_{m-1}\xi_m \prod_{l=m+1}^{t}\chi_l\right\} =$$

$$= \lim_{t\to\infty}\left\{\frac{E\,|\,\vartheta\,|^2}{(1+[t^{1-\alpha}])^2} + \frac{(t+[t^\alpha])\,\sigma_x^2\sigma_\xi^2}{(1+(t+[t^\alpha])\,\sigma_x^2)^2}\right\} = 0.$$

$$2.\quad \lim_{t\to\infty}\frac{1}{(t+[t^\alpha])^2\,\sigma_x^4}\,E\sum_{k=-[t^\alpha]+1=m}^{t}\sum^{t} x_{k-1}x_{m-1}\xi_k\xi_m =$$

$$= \lim_{t\to\infty}\frac{\sigma_\xi^2}{\sigma_x^2}\,\frac{1}{t+[t^\alpha]} = 0.$$

$$3.\quad \lim_{t\to\infty} E\left\{\frac{2}{(t+[t^\alpha])\,\sigma_x^2}\prod_{k=0}^{t}\chi_k c_0 \sum_{k=-[t^\alpha]+1}^{t} x_{k-1}\xi_k - \right.$$

$$\left. -\frac{2}{(t+[t^\alpha])\,\sigma_x^2}\sum_{k=0}^{t}\sum_{m=-[t^\alpha]+1}^{t}\varkappa_k x_{k-1}x_{m-1}\xi_k\xi_m \prod_{l=k+1}^{t}\chi_l\right\} =$$

$$= -\lim_{t\to\infty}\frac{2}{t+[t^{2\alpha-1}]+2\,[t^\alpha]}\,\frac{\sigma_\xi^2}{\sigma_x^2} = 0.$$

Следовательно,

$$\sqrt{t+[t^\alpha]}\;c_t^{(t)} \xrightarrow[t\to\infty]{\text{с. к.}} \frac{\sigma_\xi}{\sigma_x}\,\frac{S_t}{s_t}, \qquad (1.5)$$

где $S_t = \sum_t x_{k-1}\xi_k$, $s_t^2 = (t + [t^\alpha])\,\sigma_x^2\sigma_\xi^2$; $s_t^2 = ES_t^2$. Из (1.5) следует первое утверждение теоремы. Заметим, что

$$E\,|x_{t-1}\xi_t|^3 < \infty,$$

$$s_t^2 = E\,|x_{-[t^\alpha]+1}\xi_{-[t^\alpha]+2} + \ldots + x_{t-1}\xi_t|^2 = (t + [t^\alpha])\,\sigma_x^2\sigma_\xi^2 \xrightarrow[t\to\infty]{} \infty$$

и, таким образом, утверждение 2 леммы позволяет воспользоваться центральной предельной теоремой (см. 18.5.1 [17]), что непосредственно приводит к (I.4).

Приложение 2. Исходные уравнения имеют вид

$$da\,(t) = [F_{01} + F_1 a\,(t)]\,dt + G d\eta_1\,(t), \tag{2.1}$$

$$dy\,(t) = a\,(t)\,x\,(t-\tau)\,dt + B_0 d\eta\,(t), \tag{2.2}$$

где t — непрерывное время; $\eta_1\,(t)$, $\eta\,(t)$ — некоррелированные и неровские процессы; $a\,(t) = a_1\,(t)$; $\mathrm{d}y\,(t)/\mathrm{d}t \equiv x\,(t)$; F_{01}, F_1, G, B_0^2 — известны априори; $\tau = \Delta t$, Δt — интервал дискретизации при дискретном t.

Выполняя вывод уравнения для функции плотности вероятностей $\pi(\boldsymbol{a}_1^T/\boldsymbol{x}_0^T)$ аналогично [18], получаем выражение

$$\pi\,(\mathbf{a}_1^T/\mathbf{x}_0^T) = K_T\pi\,(\mathbf{a}_1^{t-1}/\mathbf{x}_0^{t-1})\,p\,(a_t/a_{t-1})\exp\left\{-\frac{1}{B_0^2}\int_{t-1}^{t}[x_t - a_t x_{t-\tau}]^2 dt\right\}\times$$
$$\times\,\pi^{-1}\,(a_{t+1})\,\pi\,(\mathbf{a}_{t+1}^T/\mathbf{x}_{t+1}^T)\,p\,(a_{t+1}/a_t)\exp\left\{-\frac{1}{B_0^2}\int_{t}^{t+1}[x_t - a_t x_{t-\tau}]^2 dt\right\}, \tag{2.3}$$

где $\boldsymbol{a}_1^T = (a_1,\ a_2,\ \ldots,\ a_T)$, $\boldsymbol{x}_0^T = (x_0,\ x_1,\ \ldots,\ x_T)$ — траектории процессов a_t, x_t на интервале $[0,\ T]$.

Интегрируя $\pi\,(\boldsymbol{a}_1^T/\boldsymbol{x}_0^T)$, приходим к искомой плотности

$$\pi_t\,(a_t/\mathbf{x}_0^T) = K_T\exp\left\{-\frac{1}{B_0^2}\int_{t-1}^{t+1}[x_t - a_t x_{t-\tau}]^2\,dt\right\}\times$$
$$\times\int_{-\infty}^{\infty}\pi_{t-1}\,(a_{t-1}/\mathbf{x}_0^{t-1})\,p\,(a_t/a_{t-1})\,da_{t-1}\int_{-\infty}^{\infty}\pi^{-1}\,(a_{t+1})\,\times$$
$$\times\,\pi_{t+1}\,(a_{t+1}/\mathbf{x}_{t+1}^T)\,p\,(a_{t+1}/a_t)\,da_{t+1}. \tag{2.4}$$

Из прямого и обратного уравнений Колмогорова имеем:

$$p\,(a_t/a_{t-1}) = \delta\,(a_t - a_{t-1}) - \Delta\frac{\partial}{\partial a_t}[(F_{01} + F_1 a_t)\,\delta\,(a_t - a_{t-1})] +$$
$$+\frac{1}{2}\Delta\frac{\partial^2}{\partial a_t^2}[G^2\delta\,(a_t - a_{t-1})] + 0\,(\Delta^2), \tag{2.5}$$

$$p(a_{t+1}/a_t) = \delta(a_{t+1} - a_t) - \Delta(F_{01} + F_1 a_t)\frac{\partial}{\partial a_t}\delta(a_{t+1} - a_t) -$$
$$- \frac{1}{2}\Delta G^2 \frac{\partial^2}{\partial a_t^2}\delta(a_{t+1} - a_t) + 0(\Delta^2). \quad (2.6)$$

Подставляя (2.5), (2.6) в (2.4) и вычисляя K_T с помощью предельного перехода $\Delta \to 0$, приходим к выражению

$$\begin{aligned} {}^+\pi_t(a_t)\,{}^-\pi_t(a_t) &= (1 - {}^+\gamma\Delta)(1 - {}^-\gamma\Delta)(1 + {}^+F(t-1, a_t)\Delta) \times \\ &\times (1 + {}^-F(t+1, a_t)\Delta)\Big\{{}^+\pi_{t-1}(a_t) - \Delta\frac{\partial}{\partial a_t} \times \\ &\times [(F_{01} + F_1 a_t)\,{}^+\pi_{t-1}(a_t)] + \frac{1}{2}\Delta\frac{\partial^2}{\partial a_t^2}[G^{2}\,{}^+\pi_{t-1}(a_t)]\Big\} \times \\ &\times \Big\{{}^-\pi_{t+1}(a_t) - \Delta(F_{01} + F_1 a_t)\frac{\partial}{\partial a_t}[{}^-\pi_{t+1}(a_t)] - \\ &- \frac{1}{2}\Delta G^2 \frac{\partial^2}{\partial a_t^2}\,{}^-\pi_{t+1}(a_t)\Big\}, \end{aligned} \quad (2.7)$$

где ${}^+\pi_t(a_t) = \pi_t(a_t/x_0^t)$; ${}^-\pi_t(a_t) = \pi_t(a_t/x_t^T)$.

Следовательно, ${}^+\pi(a_t)$ удовлетворяет обычному уравнению фильтрации, а ${}^-\pi(a_t)$ определяется выражением в обращенном времени

$$\frac{\partial\,{}^-\pi_t(a_t)}{\partial t} = -(F_{01} + F_1 a_t)\frac{\partial}{\partial a_t}\,{}^-\pi_t(a_t) - \frac{1}{2}G^2\frac{\partial^2}{\partial a_t^2}\,{}^-\pi_t(a_t) +$$
$$+ [{}^-F - \overline{({}^-F + F_1)}]\,{}^-\pi_t(a_t), \quad (2.8)$$

где ${}^+F = -(1/B_0^2)[x_t - a_t x_{t-\tau}]^2$; ${}^-F = -(1/B_0^2)[x_t - a_t x_{t+\tau}]^2$.

Выражение, связывающее $\pi(a_t/x_0^T)$ с ${}^+\pi(a_t)$, ${}^-\pi(a_t)$, получается непосредственно с помощью формулы Байеса

$$\pi(a_t/\mathrm{x}_0^T) = \frac{\pi(\mathrm{x}_0^t)\pi(\mathrm{x}_t^T)}{\pi(\mathrm{x}_0^T)} \cdot \frac{\pi(a_t/\mathrm{x}_0^t)\,\pi(a_t/\mathrm{x}_t^T)}{\pi(x_t/a_t)\,\pi(a_t)}. \quad (2.9)$$

В гауссовском приближении из (2.8) и обычного уравнения фильтрации вытекают уравнения (10), а из (2.9) — (10).

Заметим, что при $x_{t-\tau} = 1$ из полученных выражений вытекают результаты [19]. Содержание этой работы в значительной мере определило интерес к рассмотренной постановке задачи.

ЛИТЕРАТУРА

1. *С. П. Баронин.* Статистические методы анализа речевых сигналов.— Электросвязь, 1966, № 5.
2. *F. Itakura, S. Saito.* A statistical method for estimation of speech spectral density and formant frequencies.— Trans. Inst. Elect. Com. Eng. Jap., 1970, v. 53-A, N 1.
3. *C.J. Gueguen, G. Carayannis.* Analyse de la parole par filtrage optimal de Kalman.— Automatisme, 1973, v. 18, N 3.

4. *T. Nakajima, H. Omura, K. Tanaka, Ichiraki.* Estimation of vocal tract area function by adaptive inverse filtering method.— Bull. Electrotechn. Lab., 1973, v. 37, N 4.
5. *B. S. Atal, M. R. Schroeder.* Adaptive predictive coding of speech signals.— BSTJ, 1970, v. 49.
6. *С. П. Баронин.* Рекуррентные методы анализа речевых сигналов.— 8-я Всесоюзная акустическая конференция (рефераты), т. 1. М., 1973.
7. *Н. Н. Акинфиев, Р. Г. Загубная.* Адаптивное устройство для детектирования текущих параметров линейной прогнозирующей системы.— АРСО-8. Ч. 1. Львов, 1974.
8. *М. В. Назаров, Ю. Н. Прохоров.* К вопросу о методах сокращенного представления речевых сигналов.— Труды учебных ин-тов связи, вып. 71, 1975.
9. *J. N. Maksym.* Real-time pitch extraction by adaptive prediction of the speech waveform.— IEEE Trans, 1973, v. AU-21, N 3.
10. *J. D. Markel, A. H. Gray.* A linear prediction vocoder simulation upon the autocorrelation method.— IEEE Trans., 1974, v. ASSP-22, N 2.
11. *C. Scagliola.* Automatic vocal tract parameter estimation by an iterative algorithms.— C SFLT Rapporti techici, 1975, v. 3, N 2. Giugno.
12. *B. Gardini, A. Serra.* Identification of speech parameters using a recursive method.— Speech. Comm. Seminar. Stockholm, Aug. 1—3, 1974.
13. *Ю. Н. Прохоров, М. В. Назаров.* Последовательное оценивание параметров дискретного процесса авторегрессии.— 6-я конференция по теории кодирования и передачи информации, ч. 1, 1975.
14. *Ю. Н. Прохоров.* Исследование рекуррентных методов оценивания параметров сокращенного представления речевых сигналов. Автореф. канд. дис. М., 1975.
15. *М. В. Назаров, Ю. Н. Прохоров.* Оптимальное оценивание параметров речевых сигналов.— Электросвязь, 1975, № 10.
16. *Р. Ш. Липцер, А. Н. Ширяев.* Статистика случайных процессов. М., «Наука», 1974.
17. *И. А. Ибрагимов, И. А. Линник.* Независимые и стационарно связанные величины. М., «Наука», 1965.
18. *В. И. Тихонов.* Статистическая радиотехника. М., «Советское радио», 1966.
19. *С. Д. Свет.* Исследование специальных методов оптимальной фильтрации случайных процессов. Автореф. канд. дис. М., 1971.

Е. И. Кабанова

Процедуры представления сигналов, искаженных нестационарными помехами

В задачах интерпретации и распознавания обычно используется аддитивное представление исходных описаний через заданные функции. Для достижения хорошей аппроксимации эти функции должны отражать специфику описываемого процесса. В работе [1] обсуждается проблема представления описаний экспериментальных данных. Настоящая статья является развитием этой работы. Рассматривается представление экспериментальных кривых смесью экспоненциальных функций не только при стационарных, но и при нестационарных шумах записи. Используется метод «анализа через синтез».

Для исходного описания сигнала $S(x)$, измеренного в точках $x_1, x_2, \ldots, x_n$ при заданном n и заданной системе экспоненциальных функций $\varphi_1(x), \varphi_2(x), \ldots, \varphi_m(x)$, где $m \geqslant n$, строится модель сигнала

$$F(x) = \sum_{k=1}^{m} \beta_k \varphi_k(x).$$

Число ненулевых компонент невелико (2 или 3) и зависит от структуры исходных описаний. Параметры $\beta_1, \beta_2, \ldots, \beta_m$ образуют представление b, выходной вектор процесса. Представление определяется рекуррентно

$$\overset{\tau}{b} = E(S, F(\overset{\tau-1}{b}), \tau),$$

где E — правило последовательного уточнения представления b; τ — время процесса переработки описания в представление, номер шага. Правило определяется сравнением S и F и оптимизацией F по некоторому показателю качества Q. Предполагается [2], что функция Q известна с точностью до параметров b, которые должны быть определены в процессе переработки. Таким образом, задача сводится к определению конечного числа параметров $\beta_1, \beta_2, \ldots, \beta_m$. В рассматриваемой задаче этими параметрами являются ненулевые интенсивности и соответствующие им постоянные времени экспонент.

Один из способов решения задачи многопараметрической оптимизации — однопараметрический поиск по частным показателям качества (метод Гаусса — Зайделя). Он состоит в циклическом чередовании оптимизаций по всем управляемым параметрам. Для решения задачи используется один частный показатель качества, например квадратичный: $Q_j = \sum_{i=1}^{n} [S(x) - F(x)]^2$.

Координатор J задает порядок обращения к управляемым параметрам. На каждом этапе поиска происходит минимизация показателя Q_j по рассматриваемому параметру, и полученное значение параметра вводится в модель сигнала F.

Применение метода [2] удобно, если параметры функции качества Q независимы. Отсутствие перекрестного влияния параметров дает возможность в процессе оптимизации обращаться к каждому параметру лишь один раз, так как для такой функции положение экстремума по каждому параметру не зависит от значений других параметров. Поэтому задача решается за один цикл, и результат не зависит от порядка оптимизации параметров.

В работе [1] исходное описание представляет собой смесь трех экспонент с равномерным шумом. Относительная независимость оптимизаций по частным показателям качества обеспечивается фильтрацией описания и его модели режекторными фильтрами

$$R = \sum_{k=0}^{l} a_k E^k,$$

где $E^k y(x) = y(x + k\Delta x)$. Простейший режектор не пропускает экспоненциальный сигнал, на постоянную времени которого он настроен. На каждом этапе оптимизации осуществляется фильтрация сигнала двумя режекторами, настроенными на компоненты, которые не рассматриваются на данном этапе. Выделенная компонента оптимизируется независимо по постоянной времени и по интенсивности.

Настройка режекторов производится по приближенным значениям параметров, поэтому за один цикл процесс не сходится. На следующем цикле настройка режекторов корректируется, и после оптимизации модель сигнала становится более точной. Процесс сходится за 3—4 цикла. Недостатком использования режекторных фильтров является уменьшение интенсивности выделяемой компоненты по формуле $A_R = A_1 (\exp(-\beta_1\Delta x) - \exp(-\beta_2\Delta x)) \cdot (\exp(-\beta_1\Delta x) - \exp(-\beta_3\Delta x))$, где A_1, β_1 — параметры выделяемой компоненты; β_2, β_3 — постоянные времени подавляемых компонент; A_R — интенсивность сигнала на выходе фильтра.

При сближении постоянных времени различных компонент отношение сигнал/шум резко ухудшается. При малых интенсивностях сигнала шумы в основном определяются флуктуациями в электрических цепях измерительного устройства [3].

При средних и больших интенсивностях возрастает влияние шумов, связанных с наложением импульсов и неполным собиранием заряда. Шумы оказываются нестационарными, причем наибольшая интенсивность шума наблюдается на начальном участке записи. Исходное описание представляет собой смесь двух экспоненциальных и одной постоянной компоненты. Из-за шума на начальном участке записи не удается достаточно хорошо оценить параметры быстрой экспоненты после подавления медленной режекторным фильтром. Настройки режекторов оказываются слишком грубыми, процесс расходится.

Сходимость обеспечивается членением записи на участки, в каждом из которых доминирует одна компонента. Это участок записи, на котором интенсивность рассматриваемой компоненты значительно превосходит интенсивность других компонент. На таком участке можно оценить оба параметра экспоненты, осуществляя оптимизацию по одному показателю качества. Понизив таким образом размерность задачи, удается сохранить независимость оптимизаций по частным показателям качества. Параметры, определяющие членение записи, задаются.

Вычисляются приближенные значения параметров компонент. По этим оценкам формируется модель сигнала. Осуществляется циклическая оптимизация показателей качества на соответствующих участках записи. Если приближенные значения достаточно близки к истинным, то процесс сходится, несмотря на помехи. Если же параметры членения выбраны неудачно, то оценки начального приближения грубы, и процесс переработки описания расходится. В программе, реализующей описанный алгоритм, организовано

полуадаптивное членение. Адаптивное, так как точки членения записи автоматически управляются поведением сигнала в начале и конце записи, а полуадаптивное, так как пороги, зависящие от уровня помех, вводятся как отдельные данные. Процесс оптимизации сходится за 3—5 циклов. Модель сигнала строится по формуле $F(x) = \alpha_1 \exp(-\beta_1 x) + \alpha_2 \exp(-\beta_2 x) + C$. В начале работы программы вычисляются приближенные значения параметров компонент. Приближенное значение постоянной компоненты определяется по формуле

$$C = \frac{1}{p} \sum_{x=n-p+1}^{n} S_x,$$

где $p = 20 \div 30$.

Приближенная величина интенсивности быстрой экспоненты принимается равной первому, максимальному отсчету сигнала $\alpha_1 = S_1$. Постоянная времени быстрой экспоненты оценивается по формуле

$$\beta_1^N = \sum_{x=1}^{n-N-L} (S_{x+N} - S_{x+N+L})(S_x - S_{x+L}) \Big/ \sum_{x=1}^{n-N-L} (S_x - S_{x+L})^2,$$

где $N = 2 \div 4$; $L = 2 \div 4$.

Определяется граница участка доминирования быстрой экспоненты x_B. Это число отсчетов, за которое интенсивность быстрой компоненты уменьшается в K_B раз. Затем выделяется участок, на котором доминирует медленная экспонента. Для этого сглаженный сигнал $\tilde{S}(x)$, полученный усреднением значений исходного описания по 3 соседним отсчетам, сравнивается с оценкой постоянной компоненты C. Конец участка медленной экспоненты x_M определяется из условия $\tilde{S}_{x_M} = K_M C$.

Величины K_B и K_M зависят от помех. На участке записи от x_B до x_M вычисляется постоянная времени медленной экспоненты

$$\beta_2 = \sum_{x=x_B}^{x_M-2} y_{x+1} y_{x+2} \Big/ \sum_{x=x_B}^{x_M-2} y_x y_{x+2},$$

где $y(x) = S(x) - C$.

Затем определяется приближенная величина интенсивности медленной экспоненты при $x = x_B$:

$$A_2 = \tilde{S}_{x_B} - \alpha_1 \exp(-\beta_1 (x_B - 1) \Delta x) - C.$$

Начальная интенсивность медленной экспоненты вычисляется по формуле

$$\alpha_2 = A_2 / \exp(-\beta_2 (x_B - 1) \Delta x).$$

По полученным оценкам параметров формируется модель сигнала $F(x)$ и осуществляется итеративное уточнение значений па-

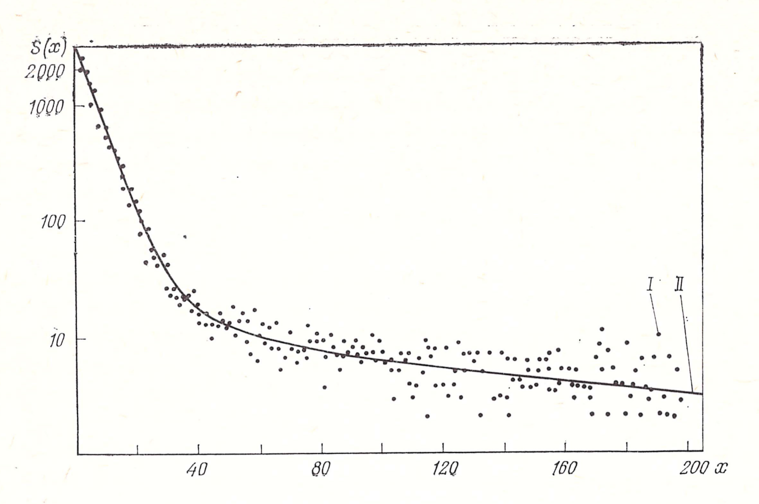


Рис. 1. Графики сигнала и аппроксимирующей функции в полулогарифмическом масштабе

I — сигнал *S* (*x*); *II* — аппроксимирующая функция *F* (*x*)

раметров по методу «анализа через синтез». Оптимизация проводится по квадратичному критерию по отдельности на участке доминирования каждой компоненты. Для быстрой, медленной и постоянной компонент соответственно:

$$Q_B = \sum_{x=1}^{x_B} (S_x - F_x)^2, \qquad Q_M = \sum_{x=x_B}^{x_M} (S_x - F_x)^2,$$

$$Q_C = \sum_{x=x_M}^{n} (S_x - F_x)^2.$$

Для каждой пары значений интенсивности и постоянной времени из набора возможных значений генерируется модель сигнала $F(x)$ и вычисляется показатель качества. Рассматриваемые параметры принимают значения, соответствующие минимуму показателя качества.

В программе используется процедура выравнивания масштабов функций S и F. Для этого в формулу вычисления функции F вводится нормирующий множитель λ: $F(x) = \lambda(\alpha_1 \exp(-\beta_1 x) + \alpha_2 \exp(-\beta_2 x) + C)$. Коэффициент λ варьируется в пределах $0{,}5 \div 1{,}5$, и для каждого значения вычисляется показатель качества

$$Q_0 = \left| \sum_{x=1}^{n} S_x^2 - \sum_{x=1}^{n} F_x^2 \right|.$$

Величины α_1, α_2 и C умножаются на коэффициент λ, соответствующий минимуму Q_0.

Ошибка аппроксимации вычисляется по формуле

$$\varepsilon = \sum_{x=1}^{n} [\ln (S_x + 1) - \ln (F_x + 1)]^2.$$

В каждом цикле переработки описания использовался следующий порядок оптимизации показателей качества: $Q_0 Q_B Q_0 Q_M Q_0 Q_C$.

На рис. 1 приводятся графики исходного сигнала и его модели, полученной в результате процесса переработки описания. Параметры модели сигнала: $\alpha_1 = 2598$; $\exp(-\beta_1 \Delta x) = 0{,}844$; $\alpha_2 = 22{,}28$; $\exp(-\beta_2 \Delta x) = 0{,}97$; $C = 2{,}57$.

Функции $S(x)$ и $F(x)$ изображены в полулогарифмическом масштабе.

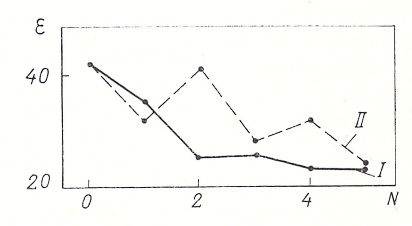


Рис. 2. Изменение ошибки аппроксимации ε для различных порядков оптимизации частных показателей качества

I — $Q_0 = Q_B = Q_0 = Q_M = Q_0 = Q_C$;
II — $Q_M = Q_0 = Q_B = Q_0 = Q_C = Q_0$;
N — номер цикла

Было проведено сравнение различных порядков оптимизации частных показателей качества. Конечные результаты переработки описания почти одинаковы, но на промежуточных этапах лучшие результаты получаются, когда вначале уточняются параметры быстрой экспоненты. На рис. 2 представлено изменение ошибки аппроксимации для различных порядков оптимизации.

Разработанный алгоритм предполагается применить для формантного анализа речевых сигналов.

ЛИТЕРАТУРА

1. *Е. И. Кабанова, В. А. Махонин.* Представление описаний.— Сб. «Нелинейные и линейные методы в распознавании образов». М., «Наука», 1975.
2. *Л. А. Растригин.* Системы экстремального управления. М., «Наука», 1974.
3. Электронные методы ядерной физики. Под ред. Л. А. Маталина. М., Атомиздат, 1973.
4. *В. В. Подиновский, В. М. Гаврилов.* Оптимизация по последовательно применяемым критериям. М., «Советское радио», 1975.
5. *Е. И. Кабанова, В. А. Махонин.* Устройство для идентификации однократных временных сигналов.— Авт. свид. № 511623. Бюлл. изобрет., 1976, № 15.

М. З. Гробман, В. Й. Тумаркин

Выделение скрытых периодичностей и формантный анализ речи

Динамические характеристики речевого тракта могут быть определены с помощью параметров экспоненциальных функций, которые описывают переходные и установившиеся процессы в нем. Измерению этих параметров должен предшествовать выбор модели анализируемого процесса, согласно которой выбирается алгоритм измерительной процедуры. Несоответствие между процессом и приписанной ему моделью может привести к значительным погрешностям.

По акустической теории речеобразования [1] наиболее адекватным физической стороне процесса является описание вокализованных участков речи экспоненциальными функциями с комплексными амплитудами и частотами. На каждый импульс голосовых связок речеобразующий тракт откликается совокупностью затухающих колебаний. Частота каждого колебания ω соответствует резонансной (формантной) частоте артикуляционного аппарата. Скорость затухания определяется добротностью резонанса. Таким образом, на интервалах времени, когда голосовые связки сомкнуты, речевой сигнал для вокализованных звуков можно представить в следующем виде:

$$y(t) = \sum_{j=1}^{N} A_j \exp(-\alpha_j t) \cos(\omega_j t + \varphi_j), \qquad (1)$$

где A_j, α_j, φ_j — соответственно амплитуда, коэффициент затухания и фаза j-й форманты. При этом в общем случае ω_j не кратны друг другу.

Отыскание этих величин представляет собой задачу, обобщающую определение характеристик скрытых периодических процессов без затухания (периодичностей). Следует отметить, что выявление скрытых периодичностей [2], т. е. распознавание спектральной структуры процессов по результатам их непосредственных измерений, отлично от разложения функции в ряд Фурье, при котором она считается периодической с периодом, равным интервалу наблюдений. Исследователь, как правило, не может достаточно полно связать интервал наблюдения с анализируемым процессом, и поэтому Фурье-анализ не позволяет достаточно точно изучить процесс.

Действительно, при таком анализе речевого сигнала форманты проявляются в виде спектральных максимумов, которые могут сливаться в один вследствие близости формантных частот и затухающего характера колебаний [3].

В методах вскрытия периодичностей периоды не навязываются заранее, а определяются в процессе самого исследования. Од-

нако применение этих методов для анализа речевого сигнала связано с рядом затруднений. Интервал наблюдения в данном случае определяется периодом свободных колебаний в речевом тракте (голосовые связки при этом сомкнуты), который для высоких голосов может составить время меньше минимального возможного периода колебаний первой форманты. Более того. На периоде основного тона могут наблюдаться всплески дополнительных колебаний, вызванные «дребезгом» голосовых связок. В силу этого интервал анализа должен составлять время не более 2—3 мс [4]. Способы выявления скрытых периодичностей, достаточно полно рассмотренные в [2], не «работают» на таких коротких интервалах времени.

В данной работе приводится описание процедуры вычисления формантных параметров речи, в основу которой положена идея разложения функции в ряд экспоненциальных составляющих [4].

Выражение (1) можно преобразовать в экспоненциальный ряд

$$y(t) = \sum_{j=1}^{N} \frac{A_j}{2} e^{-\alpha_j t} [\exp(i(\omega_j t + \varphi_j)) + \exp(-i(\omega_j t + \varphi_j))] =$$

$$= \sum_{j=1}^{N} [B_j \exp(t(-\alpha_j + i\omega_j)) + \overline{B}_j \exp(t(-\alpha_j - i\omega_j))], \qquad (2)$$

где $B_j = \frac{A_j}{2} e^{i\varphi_j}$.

Пусть значения функции $y(t)$ заданы в моменты времени t_0, $t_0 + h, \ldots, t_0 + ph$ и равны $y_0, y_1, \ldots, y_p$. Введение обозначений

$$S_{2j-1} = \overline{S}_{2j} = B_j \exp(t_0(-\alpha_j + i\omega_j)), \qquad (3)$$

$$(j = 1, 2, \ldots, N)$$

$$z_{2j-1} = \overline{z}_{2j} = \exp(h(-\alpha_j + i\omega_j)) \qquad (4)$$

позволяет, подставляя в (2) значения y_k $(k = 0, 1, \ldots, p)$, получить следующую систему уравнений относительно неизвестных S и z:

$$Y_k = \sum_{j=1}^{2N} S_j z_j^k. \qquad (5)$$

Представляя все z_j в виде корней полинома степени $2N$ с действительными коэффициентами

$$z^{2N} + \sum_{j=0}^{2N-1} b_{j+1} z^j = 0, \qquad (6)$$

можно путем несложных преобразований системы (5) получить систему линейных уравнений относительно $b_1, b_2, \ldots, b_{2N}$:

$$\sum_{j=1}^{2N} y_{m+j} b_j = -y_{m+2N+1}, \qquad m = -1, 0, 1, \ldots, p - 2N - 1. \qquad (7)$$

Очевидно, что количество точек, в которых задается функция $y(t)$, должно удовлетворять условию $p \geqslant 4N - 1$.

Система (7) решается методом наименьших квадратов [1], после чего находятся корни полинома (6). Частоты ω_j и декременты α_j находятся из (4) по формулам

$$\omega_j = \begin{cases} \frac{1}{h} \operatorname{arc\,tg} \frac{\operatorname{Im} z_{2j-1}}{\operatorname{Re} z_{2j-1}}, & \text{если} \quad \operatorname{Re} z_{2j-1} \neq 0, \\ -\pi/2h, & \text{если} \quad \operatorname{Re} z_{2j-1} = 0, \end{cases} \tag{8}$$

$$\alpha_j = -\frac{1}{h} \ln |z_{2j-1}|, \qquad (j = 1, 2, \ldots, N).$$

Приведенные соотношения, где из каждой пары комплексно-сопряженных корней нечетный номер имеет тот, у которого $\operatorname{Im} z > 0$, справедливы при условии $h < 1/2F_{\max}$. Для нахождения начальных амплитуд и фаз вводятся обозначения

$$S_{2j-1} = C_{2j-1} + iC_{2j}, \qquad j = 1, 2, \ldots, N. \tag{9}$$

Учитывая, что $S_{2j-1} z_{2j-1}^k + S_{2j} z_{2j}^k = 2\operatorname{Re}(S_{2j-1} z_{2j-1}^k)$, и подставляя в (5) преобразования (4) и (9), получаем систему линейных уравнений

$$y_k = 2 \sum_{j=1}^{N} \exp(-k\alpha_j h) \left[C_{2j-1} \cos(k\omega_j h) - C_{2j} \sin(k\omega_j h)\right],$$
$$k = 0, 1, \ldots, p. \tag{10}$$

Из решения этой системы методом наименьших квадратов находятся все C_j $(j = 1, 2, \ldots, 2N)$, откуда с учетом (3) и (9)

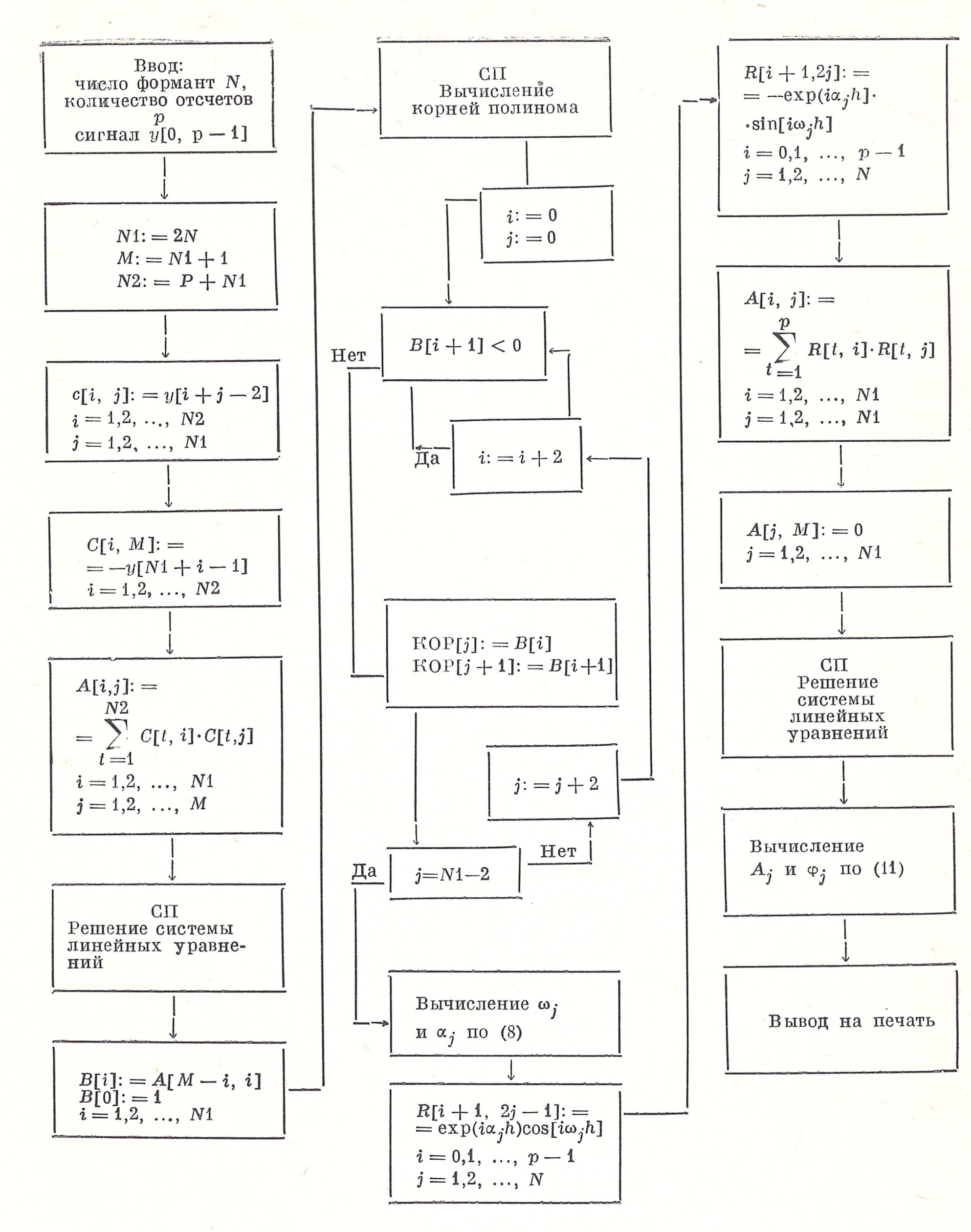
$$\varphi_j = \begin{cases} -\omega_j t_0 + \operatorname{arctg} \frac{C_{2j}}{C_{2j-1}}, & \text{если} \quad C_{2j} \geqslant 0, \quad C_{2j-1} \neq 0, \\ \frac{\pi}{2} - \omega_j t_0, & \text{если} \quad C_{2j} > 0, \quad C_{2j-1} = 0, \\ \pi - \omega_j t_0 + \operatorname{arctg} \frac{C_{2j}}{C_{2j-1}}, & \text{если} \quad C_{2j} < 0, \quad C_{2j-1} \neq 0, \\ \frac{3\pi}{2} - \omega_j t_0, & \text{если} \quad C_{2j} < 0, \quad C_{2j-1} = 0, \end{cases} \tag{11}$$

$$A_j = \frac{2\sqrt{C_{2j-1}^2 + C_{2j}^2}}{\exp(-\alpha_j t_0)} \qquad (j = 1, 2, \ldots, N).$$

Описанная процедура была реализована с помощью программы на языке АЛГОЛ-60 и отлажена на ЦВМ М-222. Блок-схема алгоритма приведена ниже.

---

[1] Матрица системы линейных уравнений, получающаяся в результате преобразования системы (7) по методу наименьших квадратов, совпадает с автокорреляционной матрицей, к которой приводит метод линейного предсказания; использование его для описания речевого сигнала предложено в [6].



Блок-схема алгоритма

C — матрица системы линейных уравнений для нахождения частот и декрементов затухания; R — матрица системы линейных уравнений для нахождения амплитуд и фаз; B — массив коэффициентов полинома; A — матрица, в которую преобразуются C и R по методу наименьших квадратов; КОР — массив корней полинома

Был проведен эксперимент на речевом материале. Параметры формант вычислялись на интервалах свободных колебаний, составляющих 2—3 мс. Полученные результаты подтвердили работоспособность процедуры.

ЛИТЕРАТУРА

1. *Г. Фант.* Акустическая теория речеобразования. М., «Наука», 1965.
2. *М. Г. Серебренников, А. П. Первозванский.* Выявление скрытых периодичностей. М., «Наука», 1965.
3. *R. W. Schafer, L. R. Rabiner.* System for automatic formant analysis of voiced speech.— JASA, 1970, v. 47, N 2.
4. *А. В. Книппер, В. А. Махонин, В. Н. Сорокин, Г. И. Цемель.* Принципы построения систем ввода речевой информации в ЭВМ.— Вопросы радиоэлектроники. Сер. ЭВТ, 1974, № 4.
5. *А. Н. Крылов.* Лекции о приближенных вычислениях. М., 1933.
6. *B. S. Atal, S. L. Hanauer.* Speech analysis and synthesis by linear prediction of the speeh wave.— JASA, 1971, v. 50, N 2.

А. В. Книппер, В. А. Махонин, И. А. Орлов

Элементы формантного анализатора

В настоящее время не существует устройств, которые могли бы измерять временные траектории формантных параметров — частоты, полосы, амплитуды и фазы формант в реальном масштабе времени. В США предпринимаются попытки создания специализированных устройств определения коэффициентов линейного предсказания, которые могут быть пересчитаны в формантные параметры [1]. Эти устройства пока сложны и дороги. Сложность измерения формантных параметров определяется сложностью и вариабильностью речевого сигнала, а подчас и неопределенностью самого понятия формант, поэтому техническая реализация точного формантного анализа, вообще говоря, имеет проблематичные акценты. Существующие методы измерения формантных частот на основе спектрального разложения [2] и измерения плотности нулевых переходов в предварительно отфильтрованных формантных полосах [3] весьма неточны. Эти неточности связаны со следующими свойствами речи:

1. Формантные области перекрываются и лежат в пределах 0,2—1 кгц для 1-й форманты; 0,5 — 2,8 кгц — для 2-й и 1,5—4 кгц — для 3-й.

2. Формантные параметры могут быстро изменяться. Так, для сочетаний носовых звуков с гласными существенные изменения происходят на одном-двух периодах основного тона, т. е. примерно в течение 10 мс.

3. Речевой процесс для вокализованных звуков на одном пе-

риоде основного тона имеет две фазы, определяемые разомкнутыми и сомкнутыми связками. Формантные параметры более точно определяются при сомкнутых связках, когда происходят свободные колебания, определяемые параметрами речевого тракта.

4. Частота основного тона и частота 1-й форманты могут быть весьма близки. Для женских голосов у некоторых звуков (например, для «и») частота основного тона может даже быть выше частоты 1-й форманты.

5. Амплитуды формант быстро уменьшаются с ростом номера форманты. Так, амплитуды 3 и 4-й формант для ряда звуков и дикторов могут быть соизмеримы с шумами придыхания, реверберации, акустическими шумами помещения, шумами квантования и схемными шумами.

Достаточно простая техническая реализация анализатора формантных параметров может быть достигнута при измерении в предварительно выделенных формантных областях. Можно отметить три основных направления увеличения точности измерения в предварительно выделенных формантных областях [4]: А. Построение фильтров с короткой импульсной реакцией, разделяющих формантные колебания. Б. Измерение формантных параметров на фазе сомкнутых связок, вызывающее необходимость построения обнаружителя моментов смыкания связок по речевому сигналу. В. Разработка метода измерения формантных параметров на коротком временно́м интервале (2—3 мс), следующим за моментом смыкания связок.

А. О разделении формантных колебаний. Как отмечалось выше, области формантных частот перекрываются. Кроме того в процессе образования речи имеют место динамические изменения формант, по [5] максимальная скорость изменения частоты 2-й форманты достигает 15 гц/мс. Все это приводит к тому, что формантные фильтры должны либо подстраиваться под траекторию движения частот формант, либо должен производиться автовыбор нескольких формантных фильтров в системе большого числа перекрывающихся фильтров. Независимо от выбранного метода импульсная реакция формантных фильтров должна быть короткой, поскольку в дальнейшем предусматривается организация измерений синхронно с основным тоном. С другой стороны, важно иметь хорошие разделяющие свойства фильтров, особенно в случае отделения высших формант от низших, для которого имеется неблагоприятное соотношение амплитуд формант.

Формирование фильтра с крутыми срезами при минимальной длительности импульсной характеристики представляет сложную проблему теории цепей. При аналоговых методах обычно создают фильтры с колоколообразной формой частотной характеристики [6] или рассчитывают заданный вид временно́й характеристики при помощи линии задержки с отводами. В обоих случаях трудно добиться малого коэффициента передачи в полосе непропускания фильтра.

Ниже описывается экспериментальный способ построения фильтра с малым коэффициентом передачи в полосе непропускания и короткой импульсной реакцией фильтра. Предварительно рассчитывается и изготавливается высокоселективный n-звенный полосовой фильтр. Для любой избирательной системы существует известное соотношение неопределенности $\Delta\omega\Delta t \cong \text{const}$.

Импульсная реакция реального n-звенного фильтра (н.ч. или в.ч.) состоит из нарастающей части и спадающей части, опреде-

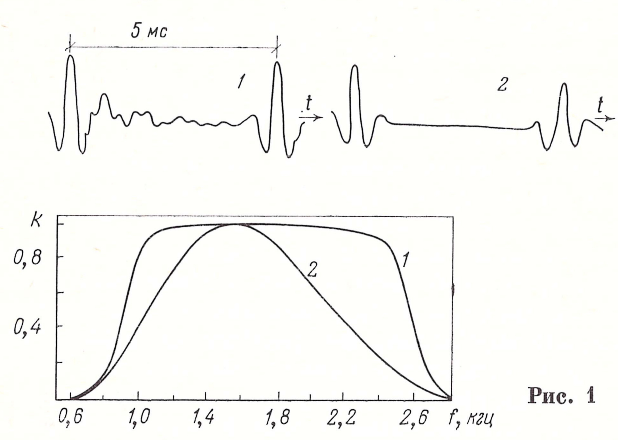


Рис. 1

ляемой в основном законом $e^{-\alpha t}\sin\omega t$. Здесь α — декремент затухания, связанный с потерями в фильтре и нагрузке, а ω — круговая частота колебательного процесса, определяемая частотами срезов фильтра. Спадающие части импульсной реакции могут быть убраны с помощью режекторов комплексных частот, включаемых последовательно с полосовым фильтром и обладающих функциями нулевой реакции на этих спадах. При этом длительность импульсной реакции Δt существенно сокращается, а срез $\Delta\omega$ — увеличивается, так что частотная характеристика полосового фильтра приобретает куполообразную форму. Эта форма частотной характеристики в полосе пропускания близка к форме ряда оптимальных спектральных окон, построенных на основе функций с двойной ортогональностью [7]. В полосе непропускания свойства фильтра определяются расчетными для полосового фильтра значениями. На рис. 1 показаны частотные и импульсные характеристики 9-звенного полосового фильтра, рассчитанного на выделение 2-й форманты. *1* — характеристики фильтра без режекторов, *2* — с режекторами, настроенными на устранение спадающих частей импульсной реакции для нижнего и верхнего срезов фильтра (временны́е характеристики скопированы с фотографии). Как видно, импульсная реакция Δt для скорректированного фильтра уменьшается в несколько раз. Эксперименты показали, что $\Delta t \cong 2{,}5\,/\sqrt{f_{н}f_{в}}$, где $f_{н}$ и $f_{в}$ — нижняя и верхняя частоты срезов фильтра.

Режекторные фильтры делаются настраиваемыми активными на RC-элементах или пассивными на LC-элементах [8]. Настройка производится по импульсной реакции фильтра регулировкой частоты и декремента затухания двух режекторов до получения максимально узкой импульсной реакции.

Б. Обнаружение моментов смыкания голосовых связок. На основе описанного формантного фильтра была разработана схема обнаружителя моментов смыкания голосовых связок. Идея работы схемы заключается в выделении формантного колебания фильтром, амплитудном детектировании этого колебания и формировании импульсов, связанных с максимумом амплитуды форманты. Как известно [8, 9], речь на вокализованных участках можно представить в первом приближении суммой равномерно затухающих синусоид — формантных колебаний. Смыкание связок вызывает амплитудный всплеск формантного колебания. Если хорошо отфильтровать соседние формантные колебания, то моменты всплеска амплитуд и будут указывать на моменты смыкания связок.

Проведенные исследования показали, что наиболее подходящей областью для выделения амплитудных всплесков является область 3-й форманты. Большинство вокализованных звуков имеет в этой области достаточно большую амплитуду форманты, а относительная высокочастотность колебания позволяет построить простой детектор формантной огибающей с малой постоянной времени, что важно для определения моментов смыкания при высоких женских голосах. Наконец, важным преимуществом схемы является возможность ее работы в спектре телефонного канала.

Блок-схема обнаружителя моментов смыкания связок проста и состоит из последовательно включенных полосового фильтра, режекторов на нижнюю и верхнюю частоты срезов фильтра, АРУ — автоматической регулировки усиления, детектора и формирователя импульсов смыкания. Формантный фильтр был выполнен на диапазон частот 1,8—4 кгц. Принципиально важной частью схемы является АРУ. Объясняется это тем, что уровень 3-й форманты для разных звуков и разных людей может меняться очень сильно — до 60 дб. Поэтому для нормальной работы детектора глубина регулировки в разработанной схеме составляет величину около 70 дб. При таких изменениях сигнала на входе АРУ сигнал на выходе изменяется не более чем на 10 дб. АРУ выполнена на двух ступенях с прямой регулировкой усиления. Детектор АРУ и детектор схемы обнаружителя импульсов смыкания Д выполнены на 4-фазной схеме. Это позволяет сделать постоянную времени сглаживающего фильтра детектора весьма малой — порядка 2 — 3 мс, что обеспечивает четкое выделение огибающей на периоде основного тона. Сглаженные в фильтре детектора сигналы огибающей имеют быстро нарастающий передний фронт и плавно спадающий задний фронт. Формирователь создает импульсы в от-

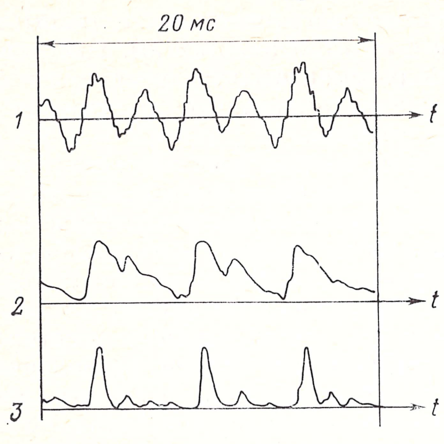


Рис. 2

вет на быстрые изменения огибающей, которые с некоторой задержкой отмечают моменты смыкания связок. Наличие глубокой АРУ вызывает срабатывание схемы не только на тональных звуках, но и на шумных. Срабатывание на тональных звуках происходит периодически, а на шумных — со случайной фазой (периодом). Схема обнаружения моментов смыкания связок, дополненная схемой выделения тона-шума, может быть положена в основу устройства измерения частоты основного тона (интонографа). На рис. 2 дан пример работы обнаружителя моментов смыкания на звуке «м» на отрезке длиной 20 мс (1 — осциллограмма звука, *2* — огибающая формантного колебания после АРУ, *3* — импульсы на выходе формирователя).

В. Синхронное измерение формантных частот. Когда формантные колебания выделены и обнаружены моменты смыкания связок, возникает задача измерения формантных параметров во временно́м окне, выбранном на наиболее благоприятном участке колебания при наиболее благоприятной длительности временно́го окна. Рядом экспериментов [4, 8] установлено, что временно́е окно должно быть выбрано в пределах 2—3 мс с небольшим смещением от максимального всплеска амплитуды формантного колебания примерно так, как это изображено на рис. 3.

Измерение частоты колебаний в таких условиях представляет существенные трудности. В известной японской системе синхронное измерение частоты формант производится по плотности нулевых переходов [3]. При этом измерение частоты 1-й форманты производится посредством отсчета времени на одном полупериоде формантного колебания, а измерение частоты 2-й форманты — отсчетом времени на семи полупериодах колебания. Такая система измерения требует разработки сложной и точной синхронизации с импульсами основного тона, наличия на колебании по крайней мере двух нулевых точек, обладает слабой защищенностью от помех, характерной для методов измерения частоты по нулевым пересечениям.

Более перспективным в этих условиях является измерение частоты на основе частотно-амплитудного метода, когда исследуемый частотно-модулированный сигнал превращается в амплитудно-модулированный, огибающая которого измеряется. Существует много разновидностей этого метода измерения частоты, мы оста-

новились на методе, описанном в [10], служащем для измерения частоты гармонического колебания.

Идея метода сводится к расщеплению формантного колебания на два колебания, отношение амплитуд которых пропорционально измеряемой частоте. Принцип работы методы измерения формантной частоты представлен на рис. 4. Измеряемый сигнал $U_{вх}$ через входной блок *Вх* поступает на дифференцирующую *Диф*

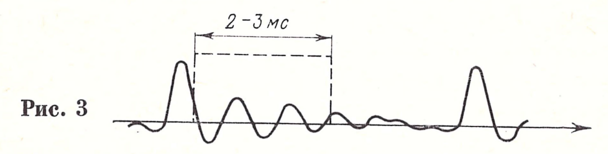


Рис. 3

и интегрирующую *Инт RC*-цепочки. Выходы с этих цепей поступают на два одинаковых канала, состоящих из детекторов *Д* и управляемых интеграторов *УИ*. Интеграторы управляются от специальной цепи, сбрасывая накопленные значения сигналов с приходом импульсов смыкания связок и работая затем в течение времени, определяемом длительностью выбранного окна. Отношение

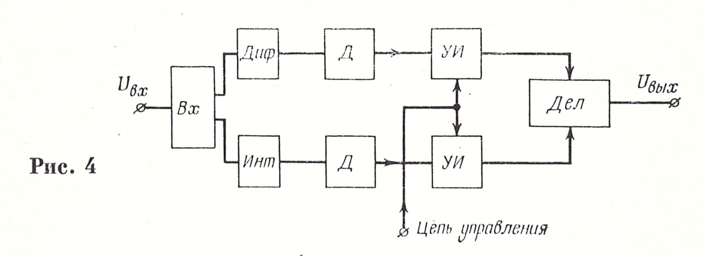


Рис. 4

сигналов в верхней и нижней частях схемы, пропорциональное измеряемой частоте, образуется в делителе напряжений *Дел*.

Рассмотрим погрешности измерения, образующиеся при подаче на вход схемы равномерно затухающего гармонического напряжения.

Пусть $X(t) = U_{вх}/U_0 = e^{-\alpha t}\cos\omega t$. Коэффициенты передачи дифференцирующей и интегрирующей *RC*-цепей имеют соответственно вид $K_д(p) = p/(p+a)$ и $K_и(p) = a/(p+a)$, где $a = 1/RC$.

При подаче сигнала $X(t)$ на входы этих цепей получим следующие сигналы на выходах:

$$Y_д(t) = \frac{\sqrt{\alpha^2+\omega^2}}{\sqrt{(a-\alpha)^2+\omega^2}} e^{-\alpha t}\cos(\varphi(\omega) - \operatorname{arctg}\omega/\alpha) + \Delta Y(t),$$

$$Y_и(t) = \frac{a}{\sqrt{(a-\alpha)^2+\omega^2}} e^{-\alpha t}\cos\varphi(\omega) - \Delta Y(t).$$

При этом $\Delta Y(t) = a(a-\alpha)e^{-at}/[(a-\alpha^2)+\omega^2]$, $\varphi(\omega) = \omega t - \operatorname{arctg}[\omega/(a-\alpha)]$.

Анализ величины $\Delta Y(t)$ показал, что при наиболее тяжелых условиях измерения (для наиболее низкой частоты первой форманты и малой добротности) она может быть сделана падающей до пренебрежимо малой величины за время, на порядок меньшее длительности наиболее короткого периода основного тона.

Аргументы функций $Y_{\text{д}}(t)$ и $Y_{\text{и}}(t)$ отличаются на угол $\psi = \operatorname{arc\,tg}(\omega/\alpha)$. Для больших отношений ω/α этот относительный сдвиг фаз равен 90°. Для наиболее тяжелых условий измерения частоты 1-й форманты при добротности $Q = 3$ относительный сдвиг по фазе оказывается равным 81°, т. е. достаточно близким к 90°. При компенсации этого сдвига фазы во входном каскаде отношение напряжений на выходах дифференцирующей и интегрирующей цепей определяется величиной $\eta = \sqrt{\alpha^2 + \omega^2}/a$.

При измерении в наиболее тяжелом случае для 1-й форманты при $Q = 3$ погрешность измерения частоты оказывается менее 1%.

Как видно из схемы рис. 4, измерение частоты производится не на амплитудных значениях напряжений, а на усредненных. Переход к средним значениям напряжений имеет два преимущества. Во-первых, упрощается схема делителя *Дел*, поскольку он работает в этом случае от униполярных напряжений. Во-вторых, использование интеграторов при вычислении усредненных напряжений уменьшает влияние случайных помех и остатков других формант на точность измерения. Схема была реализована на полупроводниках и показала точность измерения частоты в диапазоне 1-й форманты порядка 5% при длительности временного окна 2 мс.

ЛИТЕРАТУРА

1. *J. R. Welch, J. D. Oetting.* Formant extraction hardware using adaptive linear predictive coding.— IEEE 1973. National Telecommunications Conference, 1973, v. 2. Atlanta.
2. *J. L. Flanagan.* Automatic extraction of formant frequences from continuous speech.— JASA, 1956, v. 28, N 1.
3. *K. Shirai, H. Fujisava, Y. Sakai.* Ear and voice of the WOBOT.— Bull. Sci. and Engng. Res. Lab., 1973, N 62.
4. *А. В. Книппер, В. А. Махонин.* К описанию речевых сигналов.— Сб. «Речевое общение в автоматизированных системах». М., «Наука», 1975.
5. *J. D. Broad.* Formants in automatic speech recognition.— Internat. J. Man-Machine Studies, 1972, v. 4.
6. *А. А. Харкевич.* Спектры и анализ. М., Физматгиз, 1962.
7. Функции с двойной ортогональностью в радиоэлектронике и оптике. США, 1961—1968. М., «Советское радио», 1971.
8. *А. П. Вайншток, А. В. Книппер, В. А. Махонин, И. Т. Турбович* Анализ речи режекторными фильтрами.— Сб. «Речевое общение в автоматизированных системах». М., «Наука», 1975.
9. *Г. Фант.* Акустическая теория речеобразования. М., «Наука», 1964.
10. *Я. А. Яснов.* Анализ контура самонастройки в самонастраивающейся системе автоматического управления летательными аппаратами.— Сб. «Теория и применение самонастраивающихся систем». Киев, «Техника», 1968.

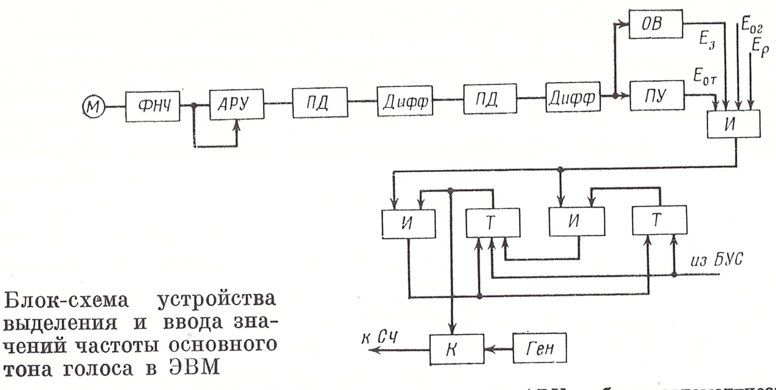
В. П. Савельев

Устройство выделения и ввода значений частоты основного тона голоса в ЭВМ

Для автоматического выделения ударных гласных при распознавании слитной речи в системах понимания речи и членения с их помощью непрерывного речевого сигнала на слова, а также для получения другой просодической информации необходимы измерения частоты основного тона голоса. Устройство выделения частоты основного тона (ОТ) F_0 должно быть достаточно простым и допускать совместную работу с цепями выделения других признаков в устройствах выделения и ввода речевых признаков (УРВ) в ЭВМ [1—3]. В связи с последним требованием устройство выделения и ввода частоты основного тона должно иметь цифровой выход с передачей отсчетов со значениями ОТ каждые 20 мс, а также должно управляться блоком управления выделением признаков и блоком синхронизации и связи с ЭВМ, имеющимися в УРВ.

Существует несколько способов выделения F_0, дающих различные результаты по точности. В данном случае требования к точности выделения F_0 менее строгие, чем, например, в вокодерной технике, так как имеется возможность программной коррекции отдельных ошибок. От автокорреляционного способа выделения F_0 пришлось отказаться ввиду его громоздкости. Методы частотной и амплитудной фильтрации частоты ОТ, описанные во многих работах (см., например, [4—7]), приблизительно равноценны, но первый несколько сложнее (необходимость подбора фазовых характеристик фильтров, значительное количество ошибок при значениях F_0, близких к граничным частотам фильтров).

Выделение F_0 было осуществлено по известной схеме: фильтрация нижних частот — нормализация максимальных пиковых значений — детектирование. Принято считать, что детектирование (сигнал без подчеркивания нижних частот) должно быть шестикратным. В нашем случае используется двойное детектирование и дифференцирование. Как правило, указанная схема амплитудной селекции частоты ОТ реализуется на двух каналах [8], обрабатывающих положительную и отрицательную полуволны речевого сигнала. Поскольку знак асимметрии речевого потока в зависимости от произнесения различными дикторами практически не меняется, была использована одноканальная схема выделения F_0 (фаза цепи: «микрофон — устройство выделения F_0''» постоянна), но при определенных ограничениях. Так, диапазон выделяемых частот ОТ был ограничен частотами 90—400 гц, что, согласно статистическим исследованиям, справедливо в 95% случаев. Кроме того частоты ОТ ниже 90 гц не могли быть введены в ЭВМ при выбранном способе ввода периода ОТ (см. ниже), значения которого вводятся в ЭВМ через $\Delta t = 20$ мс, так как внутри этого интервала должно быть не менее двух периодов ОТ.



Блок-схема устройства выделения и ввода значений частоты основного тона голоса в ЭВМ

М — микрофон; ФНЧ — фильтр нижних частот; АРУ — блок автоматического регулирования уровня; ПД — пиковый детектор; Дифф — дифференцирующая цепь; ОВ — одновибратор; ПУ — пороговое устройство; И — блок логического умножения; Т — статический триггер; К — ключ; Ген — генератор частоты заполнения периода ОТ; Сч — счетчик; БУС — блок управления и синхронизации с ЭВМ

Указанное выше ограничение рабочего диапазона выделяемых частот ОТ позволило усилить низкочастотное подчеркивание речевого сигнала. Спад частотной характеристики фильтра нижних частот (ФНЧ) вместо обычных 6 дб на октаву доведен до 12 дб. Воздушная струя при произнесении взрывных звуков (особенно для «п» и «т») вызывает в таком фильтре мощный нестационарный процесс, что часто приводит к ошибкам измерения первых периодов ОТ последующих гласных звуков. Поэтому перед микрофоном должен стоять экран (диаметром около 2 см), конструктивно объединенный с ограничителем расстояния до губ диктора (около 6 см). Такая близость объясняется необходимостью увеличения отношения сигнал/шум в условиях работы УРВ в машинных залах ЭВМ. Так как громкость произнесения должна находиться в определенных пределах, то после фильтрации диапазон изменения речевого сигнала составляет не более 40 дб.

Для нормализации уровня после фильтрации на выходе ФНЧ используется АРУ с прямой регулировкой. Для мгновенного компрессирования время заряда емкости на выходе пикового детектора АРУ выбрано достаточно малым. Скорость же разряда должна превышать максимальную скорость изменения огибающей сигнала на участках гласных, но быть не менее некоторой величины, иначе мешающие колебания между максимальными пиковыми значениями сигнала будут относительно увеличиваться, что несколько скомпенсирует низкочастотное подчеркивание сигнала, произведенное в фильтре нижних частот. Экспериментальным путем постоянная времени фильтра детектора АРУ была установлена равной 10 мс.

Следующий элемент устройства выделения частоты ОТ — схема выделения пиковых значений сигнала имеет меньшую, чем

обычно, постоянную времени фильтра, но детектор работает с отсечкой около 30%. Это позволяет более эффективно срезать мешающие выбросы на сигнале, приводящие к ложным отсчетам F_0, при меньших колебаниях выходного сигнала пикового детектора в рабочем диапазоне. Колебания выходного сигнала связаны с тем, что изменения огибающей пиковых значений сигнала на выходе АРУ увеличиваются с частотой и тем сильнее, чем больше постоянная времени разряда конденсатора на выходе пикового детектора. Это затрудняет установку порога реле.

На рисунке дана блок-схема устройства выделения и ввода в ЭВМ «Минск-22» частоты основного тона. Фактически выделяется период основного тона. На выходе устройства период ОТ представлен некоторым числом импульсов частоты заполнения этого периода ($f = 7{,}8$ гц). Эта частота выбрана, исходя из следующего. С одной стороны, при низшей F_0 ее значение не должно занимать более 7 двоичных разрядов в памяти ЭВМ, с другой — при высоких значениях F_0 не должны теряться существенные для выделения ударения, изменения F_0 на верхнем пределе 400 гц. На интервале 20 мс измеряется только первый период ОТ после предыдущего отсчета (блоки *И* и *Т*).

Предварительные эксперименты показали, что, несмотря на отсутствие второго канала, надежность выделения частоты основного тона составляет не менее 92%. Если отбросить первый, часто неверный отсчет, то надежность выделения частоты ОТ повышается до 97%. Значительную часть ошибок (удвоение F_0) дают звуки «о» и «у» (при некоторых тональностях), но они редко встречаются на двух и более отсчетах подряд и поэтому могут корректироваться программно. Значительный процент ошибок получается при произнесении «а». Это связано с двугорбостью максимальной на периоде ОТ полуволны и не может быть устранено пиковым выделителем описанного типа даже при многократной обработке. Здесь эффективнее применение двухканальной схемы выделения F_0. В нашем случае используется простая схема запрещения второго, ложного, импульса, отстоящего не далее 2 мс, что соответствует верхней граничной частоте рабочего диапазона $F_0 = 500$ гц.

Импульсы ОТ подаются на ключ *К* через схему *И*, на входы которой подаются импульсы основного тона $E_{\text{от}}$, потенциал канала шумности E_ρ, потенциал $E_{\text{ог}}$ наиболее чувствительного амплитудного дискриминатора схемы выделения и кодирования интенсивности речевого сигнала, используемого в УРВ для сегментации потока, и импульсы $E_{\text{з}}$ от схемы запрещения со временем 2—2,5 мс. Условие появления сигнала, свидетельствующего о наличии основного тона, представляется в следующем виде:

$$E = E_{\text{от}} \wedge E_{\text{ог}} \wedge \bar{E}_\rho \wedge \bar{E}_{\text{з}}.$$

В таблице приведен пример выделения частоты ОТ во фразе: «Это жирные сазаны ушли под палубу». $N_{\text{от}}$ — число импульсов

	э	т	о	ж	й́
$N_{от}$	100 74 66 77	—	04 65 63 72	—	62 56 51 57 57
F_0	136 145 132		150 154 140		170 186 186 168 168
$F_{0ср}$	137		148		173

р	н	ые
70 71	70 74	75 75 76 75 77 101 104
140 138	140 134	135 138 134 136 132 110 108
138	137	127

с	а	з	а́
—	100 102 102 110	—	100 64 61 55 51 52
	108 108 104		152 160 160 172 170
	107		160

н	ы	у	ш	л
53 56 60 65	70 74 77 77	105 107	—	72
170 160 150	134 132 132	106 105		140
165	132	105		140

й́	п	о	д	п	а́
70 64 66	—	35 70 76	—	—	100 103 103
152 150		140 134			108 108
151		137			108

л	у	б	у
110 115 121	120 120 125	—	125 141 137 140 132
100 96	96 94		98 88 85 90
98	95		90

заполнения периода ОТ в восьмеричной системе; $F_{0\,ср}$ — усредненное значение F_0 на протяжении звонкого звука; F_0 — частота основного тона. Удвоенные значения периода ОТ обнаруживались программой и уменьшались в 2 раза. Резкие уменьшения значений $N_{от}$ на одном отсчете заменялись полусуммой предыдущего и последующего отсчетов.

В приведенном примере заметно повышение F_0 на ударных гласных в слогах «жи́», «за́» и «ли́», а также понижение F_0 в конце фразы. Последний ударный гласный «а́» может быть выделен как имеющий наибольшее значение F_0 в окончании фразы. Выделение и обработка значений F_0 выполняются в реальном масштабе времени с помощью ЭВМ «Минск-22».

ЛИТЕРАТУРА

1. *А. А. Григорян, А. А. Ерошкин, С. Н. Кринов, С. С. Ракшеев, В. П. Савельев, Г. И. Цемель.* Выделение и ввод признаков речевого сигнала для систем речевого управления.— Сб. «Автоматическое распознавание слуховых образов», АРСО-6, Таллин, 1972.
2. *А. В. Васильев, С. С. Ракшеев, Т. З. Рютина, В. П. Савельев.* Система речевого ввода информации в ЭВМ М-222.— Сб. «Речевое общение в автоматических системах». М., «Наука», 1975.

3. *А. А. Алифанов, А. М. Гельфер, С. Н. Кринов, Г. И. Цемель.* Устройство ввода речевого сигнала в ЭВМ.— Сб. «Электронная техника», Сер. 9, «Автоматизированные системы управления», вып. 2, 1974.
4. *L. O. Dalansky.* Instantaneous Pitch Period Indicator.— J. Acoust. Soc. America, 1955, v. 27, N 11.
5. *W. Kallenbach.* Die Untersuchung der Sprache mit dem Tonhohenschreider.— Frequenz, 1962, Bd 16, N 2.
6. *G. U. Righini.* A pitch extractor of the voice.— Acustica, 1963, v. 13, N 4.
7. *O. O. Gruenz, L. O. Schott.* Extraction and portrayal of pitch of speech sounds.— JASA, 1949, v. 21, N 5.
8. *T. Sugimoto.* On the pitch sensation and the fundamental frequency extraction of speech sounds.— Electr. Com. Lab. Techn. J., 1962, N 2.

В. В. Максимов

Использование параллакса движения для восстановления пространственной ситуации без нахождения соответственных точек на последовательных изображениях сцены

1. Введение. Обычно считается, что использование параллакса для определения третьей пространственной координаты по двум плоским проекциям объемного мира (различают *бинокулярный параллакс*, когда сопоставляются изображения, полученные при помощи двух разнесенных в пространстве оптических систем, и *параллакс движения*, когда сопоставляются изображения, получающиеся последовательно во времени с помощью одной оптической системы в процессе ее движения в пространстве) сводится к: 1) выделению некоторых особых точек на изображениях; 2) нахождению соответствия между точками одного и другого изображения; после этого 3) простые тригонометрические расчеты дают расстояния до каждой из этих точек в пространстве. К сожалению, этот метод не лишен недостатков. Так, заранее предполагается, что форма предметов или характерная раскраска их поверхности создает на изображении какие-либо особые точки. Далее, нахождение соответственных точек на двух изображениях оказывается довольно трудоемкой процедурой. И, наконец, он дает местоположение в пространстве только этих точек и ничего не говорит о других точках. Иными словами, такой метод вместо восстановления внешней пространственной ситуации позволяет только определять «дальность» изолированных точек поверхности предметов. Желание же получить более подробную информацию о пространственной ситуации неизбежно приведет к увеличению перебора при поиске соответственных точек. В результате сцены, богатые особыми точками, становятся не под силу вычислительным машинам.

В настоящей статье разбирается принципиально отличный подход. Использование непрерывности изменений изображения спе-

ны в процессе движения оптической системы в пространстве, а также переход к рассмотрению гладких поверхностей предметов и непрерывных и дифференцируемых функций яркости позволил найти аналитическое решение задачи, не требующее ни первого этапа распознавания — выделения особых точек, ни второго этапа — нахождения соответственных точек на разных изображениях. Вследствие этого метод позволяет вычислять расстояния практически до любой точки видимых поверхностей (а исключение составляют, в частности, именно те точки, где изображение имеет особенность — скачок яркости или ее производной и пр.). Естественно, предложенный метод встречается со специфическими трудностями и имеет свои ограничения. Поэтому окончательное решение задачи восстановления внешней пространственной ситуации, по-видимому, нужно будет искать в виде некоторой комбинации этих двух методов.

2. Постановка задачи. Трехмерная сцена характеризуется, во-первых, формой и расположением объектов в пространстве и, во-вторых, распределением яркости по поверхностям этих предметов. Эти характеристики независимы в том смысле, что одна и та же по форме трехмерная сцена может быть по-разному раскрашена и освещена, и наоборот, одинаковые оптические изображения могут создаваться совершенно разными по форме сценами. При этом непосредственно может восприниматься только яркость точек поверхностей, на выходе же наша система должна давать первую характеристику сцены — объемную ситуацию.

Сцену нам будет удобно представлять в сферических (полярных) координатах, связанных с центром оптической системы. Пусть в некоторый момент времени видимая пространственная ситуация описывается расстоянием $\rho(\varphi, \vartheta)$ от центра оптической системы до видимой поверхности предмета в направлении (φ, ϑ), а функция $E(\varphi, \vartheta)$ характеризует яркость соответствующей точки поверхности. Поскольку система движется в пространстве, ρ и E будут также меняться во времени.

Итак, задача системы состоит в том, чтобы по яркости видимых точек пространства $E(\varphi, \vartheta, t)$ в каждый момент времени восстановить их удаленности $\rho(\varphi, \vartheta, t)$.

3. Упрощающие ограничения. Будем считать, что функция $E(\varphi, \vartheta, t)$ является непосредственным входом системы. Иными словами, предполагается, что оптическая система не содержит фокусирующих элементов, а изображение сцены строится проективным преобразованием (с бесконечно большой разрешающей способностью).

Трехмерная сцена представляет собой совокупность неподвижных непрозрачных предметов в прозрачной среде. Предметы ограничены кусочно гладкими поверхностями. Яркость каждой поверхности — почти везде непрерывная и дифференцируемая по поверхности функция. Яркости точек поверхностей не зависят от угла зрения.

Для простоты будем считать, что скорость системы $\bar{v}$ постоянна (причем направление $\vartheta = 0$ системы координат совпадает с направлением вектора скорости). Случай произвольного трехмерного движения не вносит ничего принципиально нового — более громоздкое аналитическое решение для этого случая разобрано нами в другой работе [1].

4. Аналитическое решение. В процессе движения оптической системы координаты каждой точки внешнего мира ведут себя так, что φ остается постоянным, а изменяются только ρ и ϑ. При этом угловая скорость движения изображения некоторой точки может быть вычислена (если в этой точке определены производные от яркости) по формуле

$$\omega = -\frac{\partial E/\partial t}{\partial E/\partial \vartheta}.$$

С другой стороны, угловая скорость относительного движения точки определяется положением этой точки в пространстве и скоростью движения системы $\omega = (v/\rho)\sin\vartheta$. Отсюда выводится формула для определения расстояния до видимых точек поверхностей предметов по их изображению

$$\rho(\varphi, \vartheta, t) = -v \sin\vartheta \frac{\partial E/\partial \vartheta}{\partial E/\partial t}. \qquad (1)$$

5. О дискретной реализации. Как видно из формулы (1), решение строится независимо для разных углов φ, т. е. задачу восстановления трехмерной пространственной ситуации по двумерному ее изображению фактически можно разбить на совокупность задач, в каждой из которых восстанавливается двумерная сцена (в координатах ρ, ϑ) по ее одномерному изображению. Такое сокращение размерности удобно как с точки зрения машинного моделирования, так и для графического представления результатов.

Необходимость дифференцирования изображения по ϑ и по t создает серьезные трудности при реализации описанного способа вычисления расстояния до произвольных точек внешнего мира. Известно, каких ухищрений требует даже более простое «выделение контуров» на изображении, полученном с помощью телевизионной камеры [2]. Одна из трудностей состоит в том, что представление непрерывного и непрерывно меняющегося во времени входного изображения на дискретном растре в дискретные моменты времени требует замены производных в формуле (1) на конечные разности. Другая, может быть, даже более серьезная трудность обусловлена высокой чувствительностью предложенного метода к шумам на изображении.

Исследование всех этих вопросов, связанных с дискретной реализацией метода, естественно проводить не с помощью реальной телевизионной аппаратуры, а с помощью генерирования изображения сцены в вычислительной машине посредством специаль-

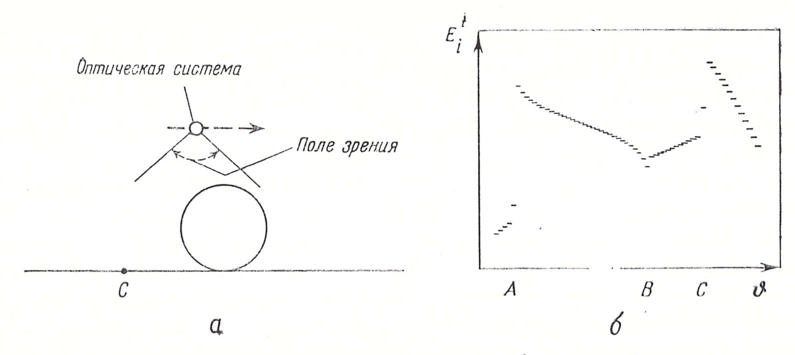


Рис. 1. Пространственная ситуация (*а*) и ее изображение (сигналы датчиков) на входе системы в момент времени $t = 75$ (*б*)

ной программы. Это позволит варьировать параметры (разрешающую способность в пространстве и во времени, число градаций яркости и шумы на входе) и исследовать влияние каждого из них в отдельности.

6. Машинное моделирование. Программа была написана на языке LEM для вычислительной машины MULTI-20 и состояла из трех блоков. С помощью одного блока в машине создавался окружающий мир — сцена (задавались форма и взаимное расположение предметов, а также яркости точек их поверхностей). Другой блок моделировал оптическую систему, формирующую изображение этого мира. Третий блок по этому изображению вычислял расстояния до видимых точек сцены.

Двумерная сцена (рис. 1, *а*) представляла собой горизонтальную прямую, на которой лежала окружность (эту сцену можно считать одним из плоских сечений некоторой трехмерной сцены, например шара, лежащего на плоскости).

Распределение яркости на предметах сцены было следующим. Яркость точек прямой слева от точки C на рис. 1, *а* линейно возрастала, в точке C претерпевала скачок, а затем (справа от точки C) убывала также по линейному закону. Яркость точек окружности изменялась как некоторая нелинейная функция дуги, для всех видимых точек окружности монотонно возрастая слева направо.

Оптическая система, которая двигалась (с заданной скоростью v) над этой сценой на некотором расстоянии от прямой (рис. 1, *а*), создавала одномерное изображение этой сцены на дискретной цепочке датчиков — «фотоэлементов». Последняя содержала 64 датчика, в сумме охватывающих поле зрения в 96° (рис. 1, *а*). Угол между оптическими осями ϑ_i соседних датчиков равнялся $\vartheta_i - \vartheta_{i-1} = \delta = 1{,}5°$. Ширина диаграммы направленности каждого из датчиков составляла также 1,5°. Выходным сигналом датчика E_i^t является интегральная яркость точек окружающего мира, проектирующихся в момент времени t на i-й датчик ($i = 0, 1, \ldots$

. . ., 63) в пределах этого угла. Пример изображения сцены, формируемого оптической системой (в один из моментов времени) представлен на рис. 1, *б*.

Вычисление расстояния $\tilde{\rho}_i^t$, где $i = 1, \ldots, 63$, до точки сцены, видимой в направлении ϑ_i, производилось по формуле

$$\tilde{\rho}_i^t = -v \frac{\sin \vartheta_{i-1}}{\delta} \cdot \frac{E_i^t - E_{i-1}^t + E_i^{t-1} - E_{i-1}^{t-1}}{E_i^t - E_i^{t-1} + E_{i-1}^t - E_{i-1}^{t-1}}. \qquad (2)$$

7. Результаты. Программа последовательно для каждого момента времени вычисляла расстояния до всех видимых точек сцены. Вычисленное программой расстояние $\tilde{\rho}_i^t$ откладывалось вдоль направления ϑ_i в полярных координатах с центром, соответствующим положению системы в момент времени t. На рис. 2 приведены отдельные кадры из построенной таким образом последовательности, начиная с того момента ($t = 25$), когда край окружности появляется в поле зрения.

В этой сцене практически везде расстояния до точек окружающего мира были вычислены с ошибками менее 0,5%. Причем основной источник ошибок — это неточное генерирование самого изображения: использование приближенных формул для интегралов, низкая точность вычисления функций $\ln x$, $\sin x$, $\cos x$ стандартными процедурами языка LEM и прочее, а совсем не замена «точной» формулы (1) на приближенную — (2). В местах изображения, где имеется скачок или излом яркости на изображении (что может соответствовать изменению яркости гладкой поверхности — точка C на рис. 1 — или границе, где один видимый предмет заслоняет другой — точки A и B на рис. 1, *б*), в двух (иногда в трех) соседних точках расстояние определяется неверно (нередко вычисленное $\tilde{\rho}_i^t$ оказывается даже отрицательным).

8. Влияние шума. Численное дифференцирование очень чувствительно к точности измерения и ко всякого рода шумам. На рис. 3, *а* показано, к чему приводит квантование яркости изображения на 256 градаций. Сравнение с тем же моментом времени $t = 75$ на рис. 2 (где яркость вычислялась с машинной точностью) показывает, что такая довольно высокая для приемного устройства точность измерения яркости для этого метода оказывается еще недостаточной.

Недостаток информации, поступающей в каждый момент времени на вход, можно компенсировать, используя априорные сведения о внешнем мире. Так, предложения о гладкости и стационарности позволяют применять обычный метод борьбы с шумами — усреднение отсчетов (или вычисление сглаженных производных по нескольким точкам). Если яркость изображения достаточно плавно меняется по пространству и во времени, то можно соответствующим образом усреднять показания соседних датчиков. Рис. 3 иллюстрирует эффект усреднения отдельно по пространству (*б*) и по времени (*в*). Это условие в действительности озна-

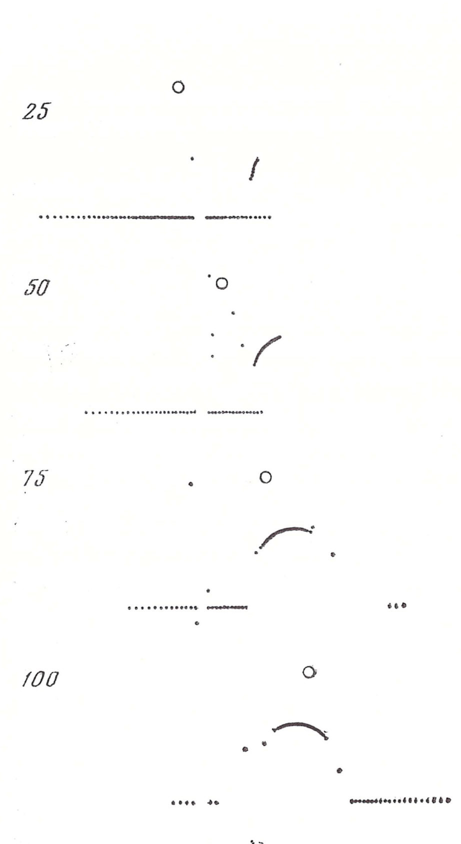


Рис. 2. Результаты вычисления расстояний до точек внешнего мира в разные моменты времени (цифры слева)

Положение системы над сценой в каждый из моментов времени отмечено кружком

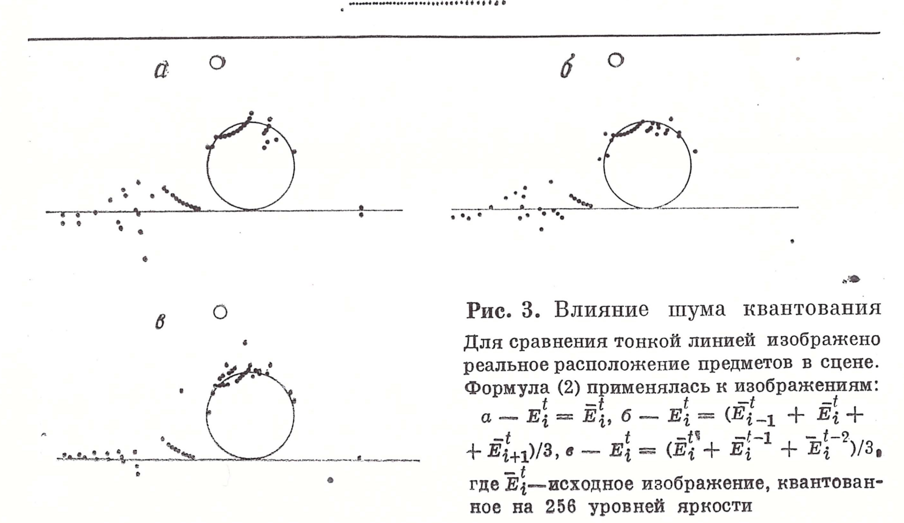


Рис. 3. Влияние шума квантования

Для сравнения тонкой линией изображено реальное расположение предметов в сцене. Формула (2) применялась к изображениям: а — $E_i^t = \bar{E}_i^t$, б — $E_i^t = (\bar{E}_{i-1}^t + \bar{E}_i^t + \bar{E}_{i+1}^t)/3$, в — $E_i^t = (\bar{E}_i^t + \bar{E}_i^{t-1} + \bar{E}_i^{t-2})/3$, где $\bar{E}_i^t$ — исходное изображение, квантованное на 256 уровней яркости

чает, что в таком мире переход к формуле в конечных разностях в принципе допустим при значительно более грубых шагах (Δt и $\Delta\vartheta$). Иными словами, метод усреднения отсчетов опирается на некоторую избыточность приемного устройства, которое при низкой точности измерения имеет (для данной сцены) чрезмерно высокую разрешающую способность. Такой размен разрешающей способности на контрастную чувствительность будет бесполезен в тех местах изображения, где усредняемая величина быстро изменяется от точки к точке.

9. Внутренняя модель внешней пространственной ситуации. В действительности, в разобранном ранее методе слабо используется априорная информация о неподвижности мира — в каждый момент времени внешняя объемная ситуация восстанавливается фактически независимо. В то же время если в некоторый момент t точно подсчитаны ρ^t, то по изображению E^t и ρ^t можно предсказать практически все изображение (за исключением изображения тех участков поверхности предметов, которые вначале находились вне поля зрения или были заслонены другими предметами) на любой момент будущего. Эту особенность можно использовать для повышения качества изображения. Поскольку каждая точка находится в поле зрения системы продолжительное время, имеется принципиальная возможность усреднять во времени отсчеты яркости отдельно для каждой видимой точки поверхности. Для этого нужно только, чтобы соответствующий накопитель перемещался по системе датчиков с той же угловой скоростью, с которой движется изображение этой точки (угловая скорость, как уже говорилось, может быть вычислена по самому изображению).

Итак, предлагается создавать в системе непрерывно улучшающуюся модель внешнего мира, которая для каждой точки сцены в каждый момент времени с большей или меньшей точностью восстанавливает: 1) ее яркость (усредняя отсчеты датчиков, на которые последовательно проектируется данная точка сцены) и 2) скорость перемещения ее изображения по системе датчиков (которая и используется для предсказания положения изображения этой точки в следующий момент). В первые моменты после того, как некоторый участок поверхности появится в поле зрения, это восстановление будет неточным. Однако с течением времени оно будет улучшаться. В результате, при не слишком больших скоростях движения система сумеет «разглядеть» (восстановить форму и яркость поверхности) каждый из предметов, появляющихся в поле зрения. Можно надеяться, что использование такой модели позволит существенно снизить требования к точности работы устройства ввода изображения, при меньшем числе априорных ограничений на воспринимаемый мир.

10. Обсуждение. Решение в конечных разностях, полученное в разделах 6 и 7, моделирует не только параллакс движения, но и бинокулярный параллакс. Следует только считать, что показания датчиков E^t и E^{t-1} с разными верхними индексами в формуле

(2) получены с помощью двух разнесенных оптических систем, а v задает расстояние между ними. Естественно, и в этом случае метод применим только к таким сценам, где производная яркости по поверхности мало меняется на расстоянии v. Поскольку в бинокулярном случае мы имеем всего два изображения внешнего мира, использование внутренней модели внешней пространственной ситуации для накопления информации теряет смысл, и единственный способ повышения точности — увеличение контрастной чувствительности самих датчиков, которое можно осуществлять, например, путем усреднения отсчетов, как это описано в разделе 8.

До сих пор при разработке систем искусственного зрительного восприятия основное внимание уделялось «упрощенному» варианту — полиэдральным сценам (составленным из многогранников). Примечательно, что предложенный здесь метод оказывается практически неприменимым к полиэдральным сценам. Действительно, с одной стороны, он не дает решения для тех точек, где производная яркости по поверхности обращается в нуль. Такая неопределенность решения может возникать на довольно больших участках поверхности, в частности на однородно освещенных (далеким точечным источником) и однородно окрашенных плоских гранях. С другой стороны, метод неприменим там, где производные не определены — в точках изображения, куда проектируются ребра многогранников и границы предметов. Только в области тени, где возможны значительные вариации освещения даже на плоских поверхностях (в результате взаимных рефлексов и взаимного затенения скученных предметов [3]), можно еще надеяться на успех.

Приверженность исследователей к полиэдральным сценам обусловлена, по-видимому, не столько тем, что такой мир является типичным, сколько надеждой на то, что результаты, полученные здесь, можно будет легко обобщить на случай криволинейных поверхностей (представляя их с большей или меньшей точностью в виде многогранников). Однако, как правило, в этих работах основным объектом анализа являются контура на изображении (соответствующие ребрам, границам предметов, границам тени и т. п.), а разрабатываемые при этом методы носят довольно громоздкий логический (переборный) характер [4], поэтому не ясно, как их можно естественно обобщить на случай произвольных сцен (когда число граней стремится к бесконечности, но при этом исчезают ребра между ними). Не удивительно поэтому, что в уже предпринятых попытках интерпретации трехмерных сцен, содержащих тела криволинейной формы, исследователи вынуждены были развивать независимый подход [5—8]. Переход к гладким поверхностям (отказ от рассмотрения ребер) в ряде случаев оказывается более продуктивным упрощением, поскольку позволяет использовать аналитический аппарат.

ЛИТЕРАТУРА

1. *В. В. Максимов, С. В. Солнцев.* Об определении расстояний до произвольных точек окружающего мира по его монокулярному изображению при известном движении наблюдателя.— Вопросы радиоэлектроники, сер. общетехн. (бионика), 1974, № 14.
2. *M. H. Hueckel.* An operator which locates edges in digitized pictures.— JACM, 1971, v. 18, N 1.
3. *В. В. Максимов, П. П. Николаев.* Цветовая оппонентность и константность цветовосприятия.— Биофизика, 1974, т. 19, № 1.
4. *P. H. Winston.* The MIT Robot.— Machine intelligence, 1972, v. 7. Univ. Press, Edinburgh.
5. *L. J. Krakauer.* Computer analysis of visual properties of curved objects.— Ph. D. Thesis, MIT, 1970.
6. *B. K. P. Horn.* Shape from shading. A method for finding the shape of a smooth opaque object from one view.— Ph. D. Thesis, MIT, 1972.
7. *G. J. Agin, T. O. Binford.* Computer description of curved objects.— Proc. Third Intern. Joint Conf. Artific. Intelligence, Stanford, 1973.
8. *И. С. Лосев, В. В. Максимов, П. П. Николаев.* Об узнавании окраски и объемной формы предметов.— Биофизика, 1975, т. 20, № 2.

Г. А. Хай

Проект диагносцирующей системы

1. Введение. Анализ диагностических врачебных ошибок показывает, что наиболее частыми их причинами являются не «упущенные» врачом, но фактически имевшие место симптомы заболевания, а трудности логической оценки и интерпретации выявленных симптомов при малом их числе. Это позволяет сформулировать в качестве центральной задачи при автоматизации диагностического процесса не выявление симптоматики, что является сегодня делом врача, а логическую обработку симптоматики [1—4]. Далее будет рассматриваться методика построения логической программы распознавания острых хирургических заболеваний органов брюшной полости.

2. Характеристика класса заболеваний. Острые хирургические заболевания органов брюшной полости характеризуются обычно внезапным началом, быстротечностью и прогрессирующим развитием, создающим непосредственную угрозу жизни больного. Значительная часть симптомов этих заболеваний встречается и при различных других заболеваниях, не относящихся к этому классу, что усугубляет возможность и опасность диагностической ошибки, поскольку при большей части острых заболеваний органов брюшной полости единственным методом лечения является неотложное хирургическое вмешательство, а при сходных по симптоматике заболеваниях, но относящихся к другим классам, обычно излечивающихся консервативными методами, операция может быть противопоказанной и даже приводить к катастрофическим последствиям.

Отнесение заболевания к выделенному классу производится при наличии любого из следующих четырех признаков: 1) жалобы на боль в животе; 2) болезненность при ощупывании живота; 3) установленный факт или подозрение на травму живота; 4) подозрение на внутреннее кровотечение. Эти данные, полученные врачом, являются основанием для формализации последующей диагностической процедуры.

3. Стратегический принцип диагностики. Из характеристики этих заболеваний видно, что практической целью их распознавания является своевременный выбор оптимального метода лечения, т. е. стратегии, от чего непосредственно зависит судьба больного. Целей академической диагностики мы здесь касаться не будем. Можно выделить следующие шесть основных стратегий хирурга: 1) показана срочная операция; 2) надо уточнить диагноз; 3) операция является напрасной; 4) операция противопоказана; 5) консервативное лечение, при неэффективности его — операция в ближайшие часы; 6) предпочтительна отсроченная или плановая операция. В соответствии с таким разбиением и перечнем стратегий хирурга целесообразен и стратегический принцип определения степени точности диагноза. Так, если число основных форм заболеваний, входящих в данный класс и распознаваемых программой, не превышает 50, то фактическое число в той или иной мере детализированных диагнозов, в том числе и динамических состояний больного, каждое из которых соответствует одной из шести основных стратегий хирурга, увеличивается более чем до 300.

4. Проявления заболеваний. Организм располагает ограниченным перечнем специфических реакций на «внутренние поломки», что определяет сходство проявления различных по содержанию патологических процессов и оказывается причиной объективных трудностей их распознавания. Таким образом, отдельно взятый симптом не может служить разумным основанием для установления диагноза болезни.

Применительно к цели данной работы под симптомами можно понимать выход за некие пороги измеримых параметров или их совокупностей, характеризующих конкретное состояние организма, причем пороги эти могут изменяться.

В процессе обследования больного врач мысленно отсеивает сотни известных ему симптомов, которых он у больного не видит, и десятки, которые он признает несущественными, оставляя для обоснования диагноза лишь 5—7 из выявленных симптомов, которые в данном сочетании оказываются наиболее содержательными. Моделируя процесс диагностики, не следует исключать те симптомы, которые могут не понадобиться в отдельном случае, но, бесспорно, могут потребоваться в любых мыслимых ситуациях. Таких симптомов для диагностики всех заболеваний выделенного класса более 1000. Они могут быть определенным образом классифицированы, так как имеют различную содержательную значимость и смысл.

5. Формализация диагноза. Синдромы. Любой патологический процесс в организме проявляет себя разнообразной симптоматикой, поскольку каждому органу свойственно множество физиологических функций, нарушения которых приводят к возникновению множества симптомов. Это дает основания рассматривать не простой перечень признаков болезни, а объединения содержательных симптомов, свидетельствующих о том или ином явлении, называемом синдромом. Такое определение синдрома здесь представляется более целесообразным, чем принятое в работах [1, 2, 5].

В соответствии с изложенным следует различать три группы синдромов: 15 *патогенетических*, 30 *дисфункциональных* и 37 *локальных*.

Симптомы, входящие в перечисленные группы, считаются основными. Кроме того, выделяется группа дополнительных симптомов, которые содержательно имеют менее общий характер, и потому они соотнесены непосредственно с конкретными заболеваниями. Часть этих симптомов выражает динамические или количественные параметры патологического процесса (нарастание, затихание; слабо, умеренно, резко выраженный основной симптом и т. д.).

Сам по себе симптом — понятие стабильное. Он может либо наблюдаться, либо отсутствовать. Напротив, синдром — понятие динамическое. Синдромы внутри каждой из трех групп могут переходить один в другой, отражая тем самым вероятные пути развития патологического процесса во времени, распространение его по различным отделам брюшной полости и вовлечение в него все более широкого круга органов и систем. При этом в случае возникновения нового синдрома предшествующий ему может сохраняться, но может и исчезать. Данные предшествовавшего наблюдения за больным и сообщаемые им о себе сведения дают основания для распознавания синдрома, отсутствующего в данный момент, но имевшего место в начальном периоде заболевания. Это дает возможность осуществить «ретроспективный поиск», имеющий существенный содержательный смысл, а также прогнозировать вероятные пути развития заболевания и возможные осложнения. Перечень таких «переходов» дополняет описание каждого синдрома.

«Эталонный образ болезни» описывается перечнем синдромов из каждой группы —«синдромным портретом». Для ряда распознаваемых состояний в этот образ включается несколько дополнительных симптомов. Естественное развитие заболевания приводит к возникновению переходов одних состояний в другие. Если при этом происходит возникновение синдрома или синдромов, не совпадающих с синдромным портретом заболевания, переход в которое данного состояния является возможным, то это свидетельствует не о реализующемся переходе заболевания, при котором может еще и не быть полного синдромного портрета, а о возникновении какой-то сопутствующей патологии, другого заболевания. Подоб-

ная структура системы распознавания дает возможность диагносцирования одновременно существующих у больного нескольких заболеваний [1]. Перечень возможных переходов дополняет описание каждого распознаваемого заболевания — его «эталонный образ».

В связи с недостаточной информацией описанный этап использования детерминистской логики может не довести распознавание до единственного диагноза. Основанием для разделения в таком случае могут служить некоторые индивидуальные характеристики больного, позволяющие считать априорную вероятность возникновения у него некоторых заболеваний (предрасположенность) весьма большой. К таким характеристикам, в частности, относятся обстоятельства, предшествовавшие развитию данного заболевания, перенесенные ранее заболевания и операции, сопутствующая патология, возраст, пол, конституция и т. п. [6, 7].

6. Описание. Принцип описания синдрома показан на следующем примере. Пусть A — синдром скрытого острого воспаления. Перечень симптомов, характеризующих данный синдром, может быть описан так: 1) пульсирующая боль; 2) постоянная боль; 3) повышение температуры тела; 4) гиперлейкоцитоз; 5) сдвиг лейкоцитарной формулы крови влево.

1-й и 2-й — взаимоисключающие симптомы, а 2, 3, 4 и 5-й могут иметь место не только в случае A, но и при трех других синдромах, описание которых мы опускаем. Для распознавания A мы пользуемся сочетаниями названных симптомов: $1 \rightarrow A$, $2 \cap 3 \rightarrow A$, $3 \cap 4 \cap 5 \rightarrow A$, $2 \cap 5 \rightarrow A$. Таким образом,

$$\{[1 \cup (3 \cap 4 \cap 5)] \underline{\cup} [2 \cap (3 \cup 5)]\} \leftrightarrow A.$$

Остальные возможные сочетания симптомов не дают клинических оснований для распознавания данного синдрома.

Аналогичным образом, но значительно проще описывается синдромный портрет заболевания — эталонный его образ. В тех случаях, когда это целесообразно по смыслу, каждому распознаваемому состоянию ставятся в соответствие те признаки, которые увеличивают у данного больного возможность данного заболевания, а также не зависящие от индивидуальных особенностей больного сравнительные частоты различных заболеваний.

7. Оптимальное решение и управление системой. Выше указывалось, что в рамках решаемой задачи установление диагноза не является самоцелью. Вместе с тем представляется желательным построение такой системы, которая в наибольшей мере соответствовала бы логике врачебного мышления. Речь идет о том, что диагноз устанавливается для выбора оптимального метода лечения больного и что, принимая решение о диагнозе, особенно о вероятностном диагнозе при недостаточной информации, врач всегда учитывает возможные последствия своего решения [1]. Поэтому «центр управления системой» должен производить сопоставление предварительных диагнозов по их стратегической

принадлежности. В случае диагностической альтернативы, но совпадения стратегий хирурга для каждого из предполагаемых заболеваний в соответствии с целью распознавания дальнейшая диагностическая процедура может быть прекращена. Продолжение ее возможно по специальному запросу врача.

При несовпадающих стратегиях процедура диагностики продолжается путем соответствующих конкретных запросов системы о недостающей симптоматике. Запрашиваются, естественно, те симптомы, которые могут дополнить имеющиеся уже симптомы для формирования эталонного образа синдрома или болезни.

Окончанием процедуры является либо однозначный диагноз, либо однозначная стратегия. При переходе к следующему этапу — вероятностному, что также может определяться «управляющим центром», решение о диагнозе будет только альтернативным. Здесь в большинстве случаев, когда одной из стратегий является хирургическое вмешательство, для выбора оптимального решения необходимо ввести в систему факторы операционного риска. Это самостоятельная программа, на которой мы не останавливаемся.

Таким образом, предполагаемая система должна состоять из трех программ — диагностики, оценки величины операционного риска и выбора оптимального решения. Врач должен иметь доступ к каждому из этапов работы системы по всем программам. Этот доступ реализуется как возможность получения решений в любом виде и как возможность введения информации на каждом из этапов в виде симптомов, синдромов, диагнозов, величины риска и прогноза. В то же время система на любом из этапов должна иметь возможность в соответствии с каждой программой запрашивать у врача недостающую ей дополнительную информацию. Таким образом, работа врача с системой должна строится в режиме диалога.

ЛИТЕРАТУРА

1. *В. А. Германов.* Дифференциальная диагностика в познании больного.— Клиническая медицина, 1969, № 9.
2. *Л. Б. Наумов.* Алгоритм диагностический.— БМЭ. Изд. 3-е. Т. 1. М., «Сов. энциклопедия», 1974.
3. *A. R. Feinstein.* An analysis of diagnostic reasoning. II The strategy of intermediate decisions.— Yale J. Biol. and Med., N 46, 1973.
4. *M. A. S'tein, J. Winter.* Theory development in medical decision-making.— Int. J. Bio-Medical Computing, N 5, 1974.
5. *А. М. Геселевич.* Термины «болезнь» и «синдром» и их определения.— Хирургия, 1971, № 4.
6. *М. Л. Быховский.* Выбор оптимального плана лечения.— Сб. «Машинная диагностика и информационный поиск в медицине». М., «Наука», 1969.
7. *Г. А. Хай.* Границы допустимого риска хирургических операций.— Экспериментальная хирургия, 1975, № 2.
8. *Л. Ластед.* Введение в проблему принятия решений в медицине. Пер. с англ. М., «Мир», 1971.
9. *Р. Ледли, Л. Ластед.* Медицинская диагностика и современные методы выбора решения. Пер. с англ. М., «Мир», 1966.
10. Распознавание образов и медицинская диагностика. Под ред. Ю. И. Неймарка. М., «Наука», 1972.

В. С. Файн

Вариативная модель в опознавании образов

В данной модели рассматриваются два процесса, с которыми связана всякая задача опознавания: процесс порождения классифицируемого ансамбля объектов и процесс собственно классификации. Оба эти процесса удобно обсудить сначала на примере конкретной задачи. С этой целью рассмотрим задачу, например, опознавания рукописных знаков.

1. Процесс порождения ансамбля. Процесс порождения некоторого знака состоит в проведении линии, связной или состоящей из нескольких частей и так или иначе изогнутой. Некоторый вариант проведения такой линии, например самый простой — близкий к прямолинейному, можно рассматривать как «первичный объект» или «праобъект». Так, на начальном этапе обучения письму детей учат именно проведению линии (писанию «палочек»), а потом уже — изменению вида этой линии.

Варьируя тем или иным способом праобъект, некоторое человеческое сообщество, объединенное общей письменностью, формирует набор объектов — алфавит используемых в данной письменности конфигураций знаков. Указанное сообщество выступает при этом в роли «генератора» или «*источника*» объектов; результат его деятельности проявляется в виде набора образцовых («эталонных») начертаний знаков — прописей.

Способы варьирования праобъекта источником объектов чрезвычайно разнообразны, но не вполне произвольны: имеется целая система очевидных физических, биологических, целевых и иных *ограничений* (система «S-ограничений»), таких, как ограничения, вносимые динамикой движения руки с пером, удобные габариты знаков, приемлемое давление пера на бумагу и др. Такие ограничения делают разные вариации праобъекта в различной степени вероятными, что позволяет обсуждать их в терминах распределения вероятностей на множестве вариаций. В рамках этого распределения выбор вариаций праобъекта, включаемых в данный алфавит, является, по-видимому, *случайным* с тем только дополнительным требованием, чтобы между знаками были не слишком малые отличия. Эта случайность выбора проявляется в данном примере в том, что биологически не отличающиеся, но разобщенные в прошлом этнически сообщества создали разные алфавиты (например, романский, грузинский, армянский, арабский и т. д.). Подобное разнообразие выборов позволяет заключить, что *природа и вид праобъекта* являются характеристическими данными для *типа источника* объектов: обсуждаемый источник, в частности, может порождать, наряду со знаками письма, самые разнообразные виды линейчатых рисунков; при этом с разными алфавитами рисунков могут быть связаны соответствующие задачи опознавания.

Таким образом, данному праобъекту (и данному типу источника) соответствует некоторый *класс* задач опознавания. Ввод той или иной системы S-ограничений и, в том числе, целевых ограничений определяет выбор из этого класса соответствующего *подкласса* задач («S-подкласса» или «S-класса») — в нашем примере подкласса задач опознавания рукописных знаков всевозможных алфавитов. Отбор же конкретного алфавита эталонных знаков и вместе с этим порождение конкретной задачи из числа задач данного S-класса осуществляется, как мы видели, случайным образом.

Рассмотрение некоторого алфавита объектов как результата варьирования (при тех или иных ограничениях) праобъекта позволяет, между прочим, заметить, что общность свойств этих объектов ничуть не менее характерна для задачи опознавания, чем их отличия, обычно больше всего интересующие разработчиков. Именно эта общность свойств, определяемая системой наложенных ограничений, обусловливает применимость для всех объектов данного алфавита общей системы описания (например, некоторой совокупности признаков), только и дающей возможность сопоставительного изучения этих объектов.

Введем для ряда употреблявшихся выше понятий формальные обозначения. Зафиксируем некоторый праобъект $\hat{A}$ и некоторый оператор $Z(\hat{A}; x_1, \ldots, x_n)$, описывающий способ варьирования праобъекта; в практически решавшейся задаче в роли Z выступали кусочные непрерывно-групповые преобразования плоскости. x_i — независимые параметры этого варьирования (его «степени свободы»). Примем для простоты, что каждый из них может принимать конечное множество значений. Множество всех возможных сочетаний значений параметров обозначим $\{X\}$; оно, очевидно, находится во взаимно-однозначном соответствии с множеством $\{A\}$ всех возможных результатов A варьирования праобъекта, которые мы будем называть его вариациями. Наложение некоторой системы S-ограничений равносильно, как было сказано, соотнесению множеству $\{X\}$ распределения вероятностей $P(X)$. В частности, наложение S-ограничений физически проявляется в том, что некоторые линии из $\{A\}$ оказываются невозможными, а некоторые другие имеют ничтожно малую вероятность появления. Иначе говоря, данная система S-ограничений выделяет из множества вариаций $\{A\}$ «практически возможное» подмножество $\{A\}_S$ (соответственно $\{X\}_S \subset \{X\}$). Множеству $\{A\}_S$ соответствует определенный класс задач опознавания — S-класс.

Зафиксируем некоторое конечное целое число $N > 0$ и, произведя N случайных выборов, выберем из $\{A\}_S$ подмножество $\{A\}_{SN}$ из N вариаций; этому подмножеству соответствует совокупность $\{X\}_{SN} \subset \{X\}_S$ из N случайных комбинаций по n чисел X_i. $\{A\}_{SN}$ образует алфавит объектов уже для конкретной задачи из S-класса. В этой задаче объекты из $\{A\}_{SN}$ могут появляться с вероятностями $P(X_j)$, $X_j \in \{X\}_{SN}$. Эти объекты были

ранее названы образцовыми или эталонными. Для краткости будем в дальнейшем называть их эталонами.

Продолжим рассмотрение примера. Люди, обучавшиеся письму по эталонным начертаниям знаков, в дальнейшем их практически не воспроизводят. Индивидуальные особенности приводят к появлению разных почерков; случайные обстоятельства, сопутствующие написанию каждого знака, создают вариативность изображений этого знака даже внутри одного почерка. Таким образом, каждое реальное начертание знака оказывается измененным по отношению к эталону. Множество всех возможных начертаний, относящихся к эталону $A_j \in \{A\}_{SN}$, обозначим $\{B\}_j$. Множество сочетаний значений параметров x_i, соответствующих всем B из $\{B\}_j$, обозначим $\{Y\}_j \subseteq \{X\}$. Вероятность возникновения начертания B, когда пишущий имел в виду j-й знак, описывается, очевидно, распределением $P(B/A_j)$ или, что то же, распределением $P(Y/X_j)$, где $Y \doteqdot B$. Объединение множеств $\{B\}_j$ для всех A_j из $\{A\}_{SN}$ образует тот ансамбль $\{B\}$ всех рукописных знаков данного языка, который является входным массивом для всякого человека, читающего произвольный рукописный текст на этом языке, или для опознающего автомата аналогичного назначения. Соответственно, в терминах параметров x_i,

$$\{Y\} = \bigcup_{j=1}^{N} \{Y\}_j.$$

Мы рассмотрели, таким образом, часть модели, относящуюся к процессу порождения ансамбля $\{B\}$ (или $\{Y\}$). Суммируем результаты этого рассмотрения.

1. Для всех объектов, фигурирующих в процессе на всех его уровнях, существует единая система описания в виде совокупностей из n величин (n разрядных «слов») X или Y.

2. Ансамбль $\{Y\}$ порождается двухступенчатым варьированием некоего исходного объекта (праобъекта), причем на каждой ступени после варьирования включается рандомизированный механизм выбора.

3. Варьирование на 1-й ступени порождает множество $\{X\}$ всех возможных вариаций праобъекта. Из этого множества случайным образом, в соответствии с распределением $P(X)$, выбирается N вариаций, образующих алфавит объектов (эталонов, образов) $\{X\}_{SN}$ в данной задаче. Если попытаться поставить в соответствие обсуждаемому процессу условную блок-схему, то 1-я ступень может быть представлена блоком, который можно назвать вариатором праобъекта или вариатором источника эталонов, выдаваемых с вероятностями $P(X_j)$.

4. Варьированию на 2-й ступени подвергаются только те N вариаций праобъекта, которые были получены по рандомизированному правилу и зафиксированы на 1-й ступени. Если на вход 2-й ступени поступает случайно выбранный из $\{X\}_{SN}$ (в соответствии с распределением $P(X_j)$) эталон X_j, то на ее выходе появляется

объект Y, получаемый случайным выбором из множества $\{Y\}_j$, в соответствии с распределением $P(Y/X_j)$. 2-й ступени в блок-схеме соответствует блок, который можно назвать «вариатором эталонов»; в памяти этого блока содержатся распределения $P(Y/X_j)$.

5. Рассмотренная часть модели состоит, таким образом, из вариаторов, в совокупности порождающих весь ансамбль возможных начертаний знаков. Это позволяет назвать указанную часть модели *вариаторной*.

2. Процесс классификации в ансамбле. Если первая часть модели описывала процесс порождения изменчивости объекта, то смысл второй части — противоположен. Здесь задача состоит в *подавлении* изменчивости, в сужении множества возможных входных (для этой части) объектов до единственного выходного объекта, индекс которого, если не произошло ошибки, должен совпасть с индексом j-го эталона, выброшенного вариатором праобъекта и затем «искаженного» вариатором эталонов.

Итак, исходным объектом в данном случае является некоторое начертание $B \in \{B\}$; ему соответствует описание $Y \in \{Y\}$. Это описание может быть порождено j-м эталоном с вероятностью $P(X_j/Y)$. Поэтому для любого Y мы должны в общем случае рассмотреть совокупность из N вероятностей $P(X_j/Y)$, учитывая при этом априорные вероятности эталонов $P(X_j)$.

Таким образом, о 2-й (собственно опознающей) части модели можно сказать следующее.

1. Эта часть имеет двухступенчатую структуру, как и часть модели, относящаяся к порождению ансамбля $\{Y\}$.

2. Первая ступень осуществляет сужение входного ансамбля $\{Y\}$ до ансамбля эталонов $\{X\}_{SN}$, представленного в этой ступени N числами $P(X_j/Y)$.

Таким образом, 1-я ступень нейтрализует изменчивость, вносимую вариатором эталонов, поэтому соответствующий блок в блок-схеме можно назвать девариатором эталонов. Для каждого $Y' \in \{Y\}$ этот блок вычисляет вероятности $P(X_j/Y')$.

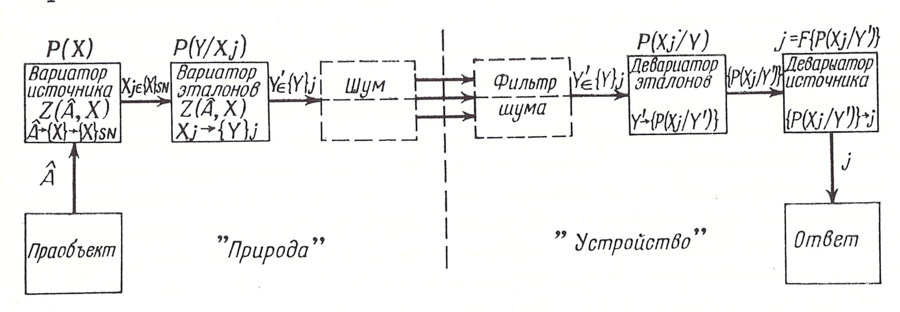
3. Вторая ступень сужает ансамбль эталонов $\{X\}_{SN}$ до единственного эталона, представленного на ее выходе номером j. Эта ступень, таким образом, нейтрализует изменчивость, создаваемую источником эталонов, и символизирующий ее в блок-схеме блок может быть назван девариатором источника.

4. В целом данная часть модели сужает весь исходный ансамбль $\{Y\}$ вариаций праобъекта до некоторой единственной вариации и потому может быть названа *девариаторной*.

3. Блок-схема модели в целом приведена на рисунке. Пунктиром показаны блоки, связанные с изменчивостью, не относящейся к природе объекта, но иногда привносимой извне в практической ситуации. Разновидностью такой дополнительной изменчивости может быть, например, шумообразное влияние шероховатости бумаги в рассмотренном примере. В связи с этим пока-

занный пунктиром блок в вариаторной части модели обозначен для простоты как «шум», а соответствующий блок в девариаторной части как «фильтр шума». Пример с шумовыми блоками подкрепляет впечатление своеобразной зеркальной симметрии модели (ее вариаторный и девариаторной частей).

4. Примеры. Приведем примеры интерпретации в терминах вариативной модели еще нескольких популярных задач.



Распознавание речевых образов. Праобъектом $\hat{A}$ может здесь считаться некоторое фиксированное состояние какого-либо одного голосового тракта, гортани и возбуждающего воздушного потока, сохраняемое неизменным в течение некоторого интервала времени T (равного длительности наиболее протяженного из распознаваемых образов, например самого длинного слова). Изменение состояния осуществляется путем сокращения мышц, управляющих формой всех частей голосового тракта, состоянием голосовых связок и диафрагмой. Если обозначить степень сокращения i-й мышцы в момент времени t через $x_i(t)$, то оператор варьирования праобъекта Z может быть записан в виде $Z(\hat{A};\ x_1(t_{11}),\ x_1(t_{12}),\ \ldots,\ x_1(t_{1m_1});\ \ldots;\ x_r(t_{r1}),\ \ldots,\ x_r(t_{rm_r}))$, где параметр $x_i(t_{ik})$ — есть число, выражающее значение функции $x_i(t)$ в k-й котельниковский момент отсчета времени. Таким образом, общее число параметров оператора Z равно

$$n = \sum_{i=1}^{r} m_i.$$

Придавая всем параметрам все возможные комбинации значений, можно получить ансамбль $\{X\}$ и одновременно $\{A\}$. Поскольку вероятности различных состояний из $\{A\}$ различны, с $\{A\}$ (и с $\{X\}$) связано распределение $P(X)$. В случае задачи опознавания речевых образов распределение $P(X)$ обусловлено S-ограничениями, выделяющими из всего множества возможных голосовых звуков (в том числе и криков, стонов, рычаний и т. д.) подмножество $\{A\}_S$ речеподобных звуков. Множество $\{A\}_{SN}$ эталонов, выбранное из $\{A\}_S$, — есть словарь данного языка, используемый в данном сообществе или в данной системе «человек — автомат». Выбором словаря $\{A\}_{SN}$ из S-класса задач рас-

познавания речеподобных образов выделяется конкретная задача. Для каждого из эталонных объектов $A_j \in \{A\}_{SN}$ существует множество $\{B\}_j$ всевозможных вариаций произнесения этого объекта разными людьми и в разных обстоятельствах; с этим множеством связано распределение $P\ (B/A_j)$ или $P\ (Y/X_j)$. Таким образом, ансамбль, с которым приходится столкнуться аудитору или опознающему автомату, а именно $\{B\} = \bigcup_{j=1}^{N} \{B\}_j$, есть в данном примере совокупность всех возможных произнесений всех N объектов (например, слов) данного словаря.

Автоматическая медицинская диагностика. Задача отличается от предыдущей тем, что здесь классифицируемый объект (структура и функции тех или иных внутренних органов и систем) непосредственно не наблюдаем, поэтому адекватного описания его изменений, в том числе и патологических, в виде оператора Z, здесь построить нельзя. В связи с этим недоступными оказываются и адекватные параметры такого оператора, и на практике в качестве параметров x_i приходится применять результаты доступных косвенных измерений и наблюдений.

Таким образом, содержательное обсуждение задачи оказывается возможным не в терминах объектов, а лишь в терминах их описаний X, Y; задача как бы начинается сразу с описания. Большая распространенность таких задач (медицинская и техническая диагностика, геофизическое прогнозирование и т. д.), надо полагать, привела к тому, что во многих работах, относящихся к опознаванию вообще, строятся и обсуждаются модели, соответствующие лишь второй, девариаторной части нашей модели; первая часть оказывается вне поля зрения.

Рассмотрим применительно к подобной задаче вариативную модель. Наибольшее и наименьшее значения всякого параметра x_i, наблюдавшиеся когда-либо на живых людях, вместе с оправдавшейся на практике дискретизацией его значений позволяют формально воспроизвести ансамбль $\{X\}$ как множество всех комбинаций возможных значений всех n параметров x_i. Эти комбинации в общем случае имеют разные вероятности. В частности, среди них могут быть и такие, при которых жизнедеятельность организма невозможна (хотя каждый отдельный параметр и может принимать данное значение на живых людях). Такова в данном случае природа S-ограничений, и с ними связано существование распределения $P\ (X)$. S-класс, отвечающий этим ограничениям, есть класс задач медицинской диагностики. В качестве описания $\hat{X}$ праобъекта можно взять, например, типичную комбинацию значений параметров для здорового человека. Выбор алфавита $\{X\}_{SN}$ определяется тем, между какими именно заболеваниями должно производиться различение; задание списка этих заболеваний выделяет из S-класса конкретную задачу.

В качестве собственно эталонов X_j из $\{X\}_{SN}$ могут выбираться типичные комбинации значений параметров для каждого из N

рассматриваемых заболеваний. Если под общим названием заболевания содержится несколько заметно отличающихся разновидностей, то в $\{X\}_{SN}$ можно ввести эталоны для каждой из них. В данном примере вариатор источника моделирует отклонения от «здорового» состояния ко всевозможным типичным болезненным состояниям, а вариатор эталонов — индивидуальные отклонения от этих типичных состояний в соответствии с распределениями $P\ (Y\ /X_j)$.

Недоступность объекта для непосредственного наблюдения в подобных задачах приводит, как отмечалось, к утрате важного конструктивного эффекта, доставляемого данной моделью,— возможности целенаправленного поиска адекватных параметров оператора Z. Это, однако, не сводит роль модели в таких задачах к чисто методологической: как будет ниже указано, принятие во внимание существования и природы множеств $\{X\}$, $\{X\}_{SN}$ и т. д. и соответствующих распределений позволяет получить важные оценки качества решения задачи, недостижимые в ином случае.

Автоматическая портретная идентификация. Праобъектом $\hat{A}$ является в этой задаче какое-либо одно человеческое лицо, например типичное лицо для данной этнической группы или данной местности. В практически решавшейся задаче в роли оператора Z выступали топологические произвольно-нелинейные преобразования поверхности лица (в частности, шаговые взвешенно-групповые преобразования) — в трехмерном варианте и такие же преобразования плоскости изображения (фотографии) — в двумерном. В последнем случае, как и при распознавании рукописных знаков, использовались также и кусочно-непрерывно-групповые преобразования. Практически такой оператор Z при варьировании его параметров x_i позволяет произвольным образом изменять как глобально, так и локально форму некоторой поверхности (в частности, поверхности лица) или форму плоского изображения с сохранением неизменной топологии поверхности или плоскости. В частности, оператор может превращать и лицо одного человека в лицо другого, и именно этот вариант его использования образует реализацию вариатора источника. S-ограничения в данном случае выделяют те из преобразований праобъекта, которые приводят к «лицеобразным» поверхностям или изображениям. С этим связан и вид распределения $P\ (A)$ или $P\ (X)$. В качестве эталонного алфавита $\{A\}_{SN}$ выступает картотека, составленная из фотографий лиц, подлежащих контролю в данной местности; априори этот набор является, очевидно, случайным. Мимические, возрастные, болезненные изменения формы лица соответствуют в нашей модели вариатору эталонов. Эти изменения описываются путем варьирования параметров качественно тех же преобразований, т. е. того же оператора Z. Возникающим при этом множествам $\{Y\}_j$ соответствуют распределения вероятностей $P\ (Y/X_j)$. Обычно на практике нет возможности изучить

для каждого из состоящих в картотеке лиц все возрастные, мимические и прочие вариации; вместе с тем биологические факторы, обусловливающие подобные вариации (места прикрепления и синергии мышц, характер увядания с возрастом кожи и т. д.), у всех людей практически одинаковы. Поэтому распределения $P(Y/X_j)$ обычно считаются для всех X_j из $\{X\}_{SN}$ одинаковыми (точнее, отличающимися только первыми моментами). Заметим, что это явление встречается во многих задачах; оно характерно, в частности, и для задач распознавания типографских, машинописных и в большой степени рукописных знаков. Одинаковость распределений $P(Y/X_j)$ вносит упрощения в вычисление распределений $P(X_i/Y)$ в девариаторной части модели.

5. Обсуждение. Отметим три основных направления, продвижению в которых может способствовать использование вариативной модели.

Первое направление связано с вводом в рассмотрение оператора $Z(\hat{X}; x_1, \ldots, x_n)$. Поскольку все члены ансамбля $\{B\}$ считаются вариациями единственного объекта — праобъекта $\hat{X}$, постольку оператор $Z(\hat{X}, x_1. \ldots, x_n)$ является описанием связей между этими вариациями, т. е. описанием связывающей их *закономерности*, а его свободные переменные x_i — параметрами, управляющими этой закономерностью. В задачах, где возможно непосредственное наблюдение объектов (как в большинстве приведенных примеров), определение физической сущности оператора Z, а затем и его формальной структуры не представляет принципиальной трудности. Но тогда автоматически оказывается решенной одна из главных проблем в задачах опознавания — проблема *выбора признаков*; в качестве признаков, как было видно, выступают параметры оператора. В частности, именно благодаря вводу в рассмотрение вариативной модели и соответствующего оператора Z удалось решить задачу автоматической портретной идентификации, до этого не поддававшуюся решению другими средствами.

Второе направление связано с неодинаковым происхождением вариаторной и девариаторной частей модели. В самом деле, изменчивость объектов, моделируемая вариаторной частью, представляет собой объективный факт внешнего мира, своего рода явление *природы*; именно таков смысл подписи под блоками вариаторов в схеме. Девариаторная же часть является моделью *устройства* или алгоритма, представляющего собой плод сознательной деятельности, направленной на подавление изменчивости. Этим обусловлена подпись под девариаторными блоками на рисунке. Природа выступает здесь, очевидно, как первичное, активное начало, порождающее изменчивость; техническое же устройство, как и имитируемый им воспринимающий аппарат живого существа, предназначено для парирования этой изменчивости, т. е. вынуждено играть пассивную, вторичную роль.

Методологический вывод из сказанного состоит в том, что изучение и описание изменчивости, т. е. ее математическое моделирование в виде оператора Z, является центральным и, по-видимому, ключевым вопросом в теории опознавания образов.

Конструктивный вывод состоит в том, что ниоткуда, вообще говоря, не следует, что изменчивость может быть *только* порождением природы. Поскольку в решении задачи опознавания мы так или иначе должны конструировать ее модель Z, постольку мы можем немедленно использовать эту модель для искусственного генерирования изменчивости. Так мы приходим к новой, отличающейся от опознания образов, области исследований — области *целенаправленного генерирования изменчивости*. Здесь имеется ряд практических задач, таких, например, как автоматизация дизайнерских работ, автоматическое построение промежуточных фаз движения в мультипликации, автоматизация некоторых работ в криминалистике. Некоторые из этих задач в настоящее время успешно решены или решаются.

Наконец, третье направление связано с создаваемой вариативной моделью возможностью перейти посредством детерминистского оператора Z от самих объектов к описаниям их вариаций в виде комбинаций из n чисел, между которыми существуют уже только вероятностные соотношения. Это позволяет ввести в рассмотрение множество всех таких комбинаций $\{X\}$ с распределением $P\ (X)$ и изучать его в сопоставлении с множествами $\{X\}_{SN}$ и $\{Y\}$ и соответствующими распределениями. Смысл перечисленных множеств и существование связанных с ними вероятностных распределений позволяет провести параллель между обсужденной выше модельной ситуацией и моделью передачи сообщений случайным n-разрядным кодом по каналу с шумами, содержащейся в известной теореме Шеннона. Эта параллель оказывается настолько глубокой, что становится возможным при некоторых дополнительных ограничениях *распространить указанную теорему Шеннона на задачи опознавания* в их вариативной трактовке. Это распространение вместе с обсуждением некоторых вытекающих из него новых возможностей в опознавании образов составляет содержание отдельной публикации.

Содержание

УДК 621.391

Об одном классе задач прогнозирования. Г и т и с В. Г. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Рассматриваются пути решения задач прогнозирования, в которых прогнозируемый показатель задается экспертом путем разбиения экспериментальной выборки на однородные и упорядоченные классы. В качестве меры соответствия прогноза и прогнозируемого показателя принимается значение условного минимума эмпирического среднего риска по порогам, задающим границы упорядоченных классов. Показано, что при введении некоторых ограничений на класс функций потерь поиск условного минимума среднего риска по порогам сводится к последовательному решению задач одномерного поиска. Библиогр. 6 назв.

УДК 621.391.199
519.25

Модифицированный метод наименьших квадратов при решении задач прогноза с дискретной прогнозируемой величиной. Ю р к о в Е. Ф., Н а г о р н о в В. С. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Формулируется критерий качества, учитывающий специфику задач с дискретной прогнозируемой величиной. Показано, что алгоритм прогнозирования, полученный в результате оптимизации введенного критерия, является модификацией метода наименьших квадратов. Библиогр. 3 назв.

УДК 621.391.19

Статистический анализ модели прогнозирования, использующей одномерные нелинейные преобразования. Ю р к о в Е. Ф. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Исследуется сложность модели прогнозирования методами проверки статистических гипотез. Приводятся экспериментальные данные, полученные в результате статистического анализа модели при решении задач медицинского и сейсмического прогнозирования. Библиогр. 7 назв.

УДК 621.391.199
519.25

Анализ связи между тяжестью состояния больного и его физиологическими показателями. Т у р б о в и ч И. Т., В и н и ц к а я Р. С., Г и т и с В. Г., Е р а м я н С. Г., Н а г о р н о в В. С., С у н г у р я н Н. Н., Ю р к о в Е. Ф. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Излагается подход к нахождению количественной связи между тяжестью состояния больного и его физиологическими показателями. В процессе исследования выбирается математическая модель связи, отбираются наиболее существенные функциональные показатели, проводится проверка возможности упрощения выбранной модели связи и производится статистический анализ данных. Приводятся экспериментальные результаты, полученные на материале 145 больных, страдающих хронической пневмонией. Библиогр. 5 назв. Иллюстраций 3.

УДК 621.391.19

Применение метода одномерных нелинейных преобразований для прогнозирования естественного режима нефтяного пласта. Б а д а л о в Т. А. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

На основании метода одномерных нелинейных преобразований решается задача прогнозирования естественного режима пласта по совокупности геолого-физических признаков, имеющихся в начальный период разработки нефтяных месторождений. Определяется характер влияния каждого из этих признаков на прогноз, а также их значимость. Используются данные, полученные на нефтяных залежах Азербайджана, разработка которых осуществлялась при очень уплотненной сетке скважин на естественном режиме. Библиогр. 7 назв. Иллюстраций 2.

УДК 621.391

К использованию алгебраических приемов идентификации систем для анализа речи. В а й н ш т о к А. П., К а б а н о в а Е. И., М а х о н и н В. А. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Описаны алгоритмы для системы идентификации параметров и состояний настраиваемой модели порождения слогов. Приведены правила оценивания отдельных параметров речеобразования. Библиогр. 10 назв. Иллюстраций 2.

УДК 621.391.534.781

Алгоритм совместного рекуррентного оценивания параметров речевых сигналов. Назаров М. В., Прохоров Ю. Н. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Методами теории нелинейной марковской фильтрации синтезирован рекуррентный алгоритм совместного слежения за динамикой модели речевого тракта и параметрами сигнала звукового возбуждения. Указаны способы упрощения схемы устройства оптимального слежения. Библиогр. 10 назв.

УДК 621.391.192.5

Элементы кодовой структуры речи. Сорокин В. Н. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Рассматривается кодовая структура речи, включающая в себя акустические и артикуляционные признаки, фонемы, слоги, слова и фразы. Вычисляется корректирующая способность фонем и слов. На основе сведений о роли фактора сложности в процессах речеобразования и восприятия, а также аналогий с теорией кодирования рассматриваются некоторые приемы анализа речевого сигнала и декодирования его фонетического состава. Установлено свойство неприводимости большинства слов, позволяющее членить слитный поток речи на слова. Приводится схема параллельного декодирования всех уровней речевого кода, в которой основным вычислительным механизмом является алгоритм последовательного декодирования. Таблиц 3. Библиогр. 24 назв. Иллюстраций 9.

УДК 621.396.64

О типологических особенностях слога. Потапова Р. К. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Показывается, что использование слога в качестве опорной элементарной структуры при автоматическом распознавании речи обусловлено прежде всего лингвистическими закономерностями. Предлагается при этом учитывать специфику слитной речи на базе конкретной языковой системы, а при анализе и описании признаков слога учитывать интегративный характер слога, как целостной структуры. Показывается, что для ряда германских языков элементарная слоговая структура не исчерпывается универсалией в рамках СГ. Подчеркивается информативность временных параметров при выявлении характера слоговой сегментации речевого континуума. Библиогр. 9 назв.

УДК 621.391:534.781

Рекуррентное оценивание параметров речевых сигналов. Прохоров Ю. Н. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Рассматриваются рекуррентные алгоритмы оценивания параметров авторегрессионной и многоэтапной моделей речевых сигналов. На основе теорий фильтрации и интерполяции марковских процессов разработаны упрощенные алгоритмы, допускающие сравнительно простую техническую реализацию. Доказана сходимость в среднеквадратическом предложенных алгоритмов и приведены результаты цифрового моделирования, выполненного на реальном сигнале. Библиогр. 19 назв. Иллюстраций 6.

УДК 621.391.199

Процедуры представления сигналов, искаженных нестационарными помехами. Кабанова Е. И. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Рассматриваются приемы представления экспериментальных кривых смесью экспоненциальных функций. Представление определяется методом «анализа через синтез». Параметры представления оптимизируются циклическим поиском по частным показателям качества. В случае равномерного шума независимость оптимизаций по частным показателям качества обеспечивается фильтрацией сигнала и его модели режекторными фильтрами. При нестационарных шумах запись расчленяется на участки доминирования отдельных компонент сигнала. Описана программа переработки описания сигнала в представление для случая нестационарных шумов. Библиогр. 5 назв. Иллюстраций 2.

УДК 621.391

Выделение скрытых периодичностей и формантный анализ речи. Гробман М. З., Тумаркин В. И. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Описывается алгоритм вычисления формантных параметров речевого сигнала, основанный на представлении анализируемых участков сигнала суммой экспоненциальных функций с комплексными амплитудами и частотами. Библиогр. 6 назв. Иллюстраций 1.

УДК 621.391.144

Элементы формантного анализатора. Книппер А. В., Махонин В. А., Орлов И. А. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Рассматривается вопрос синхронного с импульсами основного тона измерения частоты формант. Процесс измерения разбивается на три этапа: выделение формантных колебаний, обнаружение импульсов смыкания связок, измерение частоты во временнóм окне на большой длительности. Обсуждаются особенности реализации каждого из этапов. Библиогр. 10 назв. Иллюстраций 4.

УДК 621.391.193

Устройство выделения и ввода значений частоты основного тона голоса в ЭВМ. Савельев В. П. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Описывается устройство выделения частоты основного тона, удовлетворяющее требованиям совместной работы с устройством ввода речевых признаков в ЭВМ, находящимся в опытной эксплуатации. Приведены блок-схема устройства и предварительные результаты. Таблица 1. Библиогр. 8 назв. Иллюстраций 1.

УДК 681.142:155

Использование параллакса движения для восстановления пространственной ситуации без нахождения соответственных точек на последовательных изображениях сцены. Максимов В. В. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Дается аналитическое выражение, позволяющее вычислять расстояния до произвольных видимых точек внешнего объемного мира по его плоскому (меняющемуся во времени в результате движения системы) изображению. Описана дискретная реализация этого метода на вычислительной машине и приведены результаты ее работы. Метод оказывается чрезвычайно чувствительным к качеству входного изображения. Показано, что наряду с простым усреднением отсчетов по пространству и по времени для повышения качества изображения может оказаться полезным создание внутри системы модели внешнего мира (постоянно улучшающейся в процессе наблюдения). Библиогр. 8 назв. Иллюстраций 3.

УДК 621.391

Проект диагносцирующей системы. Х а й Г. А. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Рассматривается процедура распознавания острых заболеваний брюшной полости, характеризующихся началом, быстрым прогрессирующим течением и необходимостью в большинстве случаев хирургического вмешательства.

На основании симптомов, имеющихся у больного, диагносцируются три типа синдромов. Синдромы описаны в виде возможных сочетаний симптомов. Синдромный портрет — эталонный образ заболевания, описываемый перечнем синдромов.

Диагностический процесс происходит после введения симптоматики больного в систему. Если диагноз установлен, либо установлено несколько диагнозов с одной тактикой лечения, диагностика заканчивается. При несовпадающих решениях о тактике процедура диагностики продолжается путем запроса системой врача о недостающей симптоматике. Врач имеет доступ к каждому этапу работы системы, и по его желанию она может быть закончена в любое удобное для него время. Библиогр. 10 назв.

УДК 621.391.199

Вариативная модель в опознавании образов. Файн В. С. Сб. «Распознавание образов. Теория и приложения». М., «Наука», 1977.

Множество всех возможных реализаций всех классов, фигурирующих в некоторой задаче опознавания, рассматривается как результат двухэтапного варьирования какой-либо одной из реализаций с помощью специального детерминистского оператора, зависящего от конечной совокупности параметров. Из такой трактовки вытекает ряд важных следствий, в частности возможность установления аналогии с задачей передачи сообщений по шумящему каналу, внесение некоторой определенности в проблему выбора признаков и др. Иллюстраций 1.

Распознавание образов

Теория и приложения

Утверждено к печати
Институтом проблем передачи информации

Редактор издательства *Н. Н. Прокофьева*
Художник *В. В. Фирсова*
Художественный редактор *Н. Н. Власик*
Технические редакторы
В. В. Волкова, Л. Н. Золотухина
Корректоры *И. А. Талалай, В. А. Шварцер*

Сдано в набор 17/IX 1976 г.
Подписано к печати 22/XI 1976 г.
Формат 60×90$^1/_{16}$. Бумага типографская № 1
Усл. печ. л. 8. Уч.-изд. л. 8,4.
Тираж 3500. Т-16094. Тип. зак. 1228
Цена 59 коп.

Издательство «Наука»
103717 ГСП, Москва, К-62, Подсосенский пер., 21

2-я типография издательства «Наука»
121099, Москва, Г-99, Шубинский пер., 10

59 коп.